REMARKS

Claims 41-46, 50, 51 are currently pending in this application.

Applicants have amended claim 41 to include the term "transfer" before the term "vector" in the preamble and addition of the term "wherein said transfer vector transfers the defined nucleotide sequence into the nucleus of a cell." This amendment is supported, at least, on page 1. lines 2-4, on page 4, lines 3-2 and line 23, of the translated specification.

In the Final Office Action, mailed October 10, 2006, the Office maintained the rejection of claims 41-51 under 35 U.S.C. § 102 as being anticipated by either WO 97/12622 ("Verma") or Parolin et al., "Analysis in Human Immunodeficiency Virus Type 1 Vectors of *cis*-Acting Sequences That Affect Gene Transfer into Human Lymphocytes," *J. Virol.*, vol. 68, pp. 3888-95 (1994) ("Parolin"). These references, though, do not anticipate the claimed invention for the following reasons.

Verma

The Office maintained the rejection under § 102 in light of Verma because it asserted that Verma teaches "a recombinant vector comprising a polynucleotide - the HIV *pol* vector - that does include cPPT and CTS regions." Office Action mailed October 10, 2006, at page 3. Verma, though, does not teach a "transfer vector" that comprises the cPPT and CTS sequences of pol that achieves "wherein said transfer vector achieves transfer of the transgene or sequence of interest into the nucleus of a cell and forms a cis-acting triplex," as provided in the claimed invention.

The Office is apparently referring to the "first vector" depicted in page 4, lines 22-28 and Figure 1, "Vector 2" of Verma because this vector includes *pol* sequences. But,

this vector is a packaging vector in Verma, not a transfer vector. A packaging vector comprises DNA sequences encoding most structural viral proteins needed to form infectious viral particles, including *pol*. In this context, though, *pol* is incorporated into the final viral particle as a *trans*-factor and so is not able to form the DNA triplex needed for nuclear transport. Instead, the cPPT and CTS sequences of *pol* must be *cis*-acting to form a DNA triplex as provided in independent claim 41.

In Verma, the actual "transfer vector" that provides the *cis*-acting viral sequences is the "third" vector as depicted on page 6, lines 24-29, and in Figure 1 - vector 1. This vector does not include *pol* sequence, though. Furthermore, this vector is not able to form a DNA triplex.

Because Verma does not disclose a "transfer vector" capable of achieving "transfer vector achieves transfer of the transgene or sequence of interest into the nucleus of a cell and forms a cis-acting triplex" it cannot anticipate the claimed invention. Accordingly, Applicants respectfully request that the rejection under 35 U.S.C. § 102 in light of Verma be withdrawn.

Parolin

The Office maintained the rejection under § 102 in light of Parolin because it asserted that "the Parolin vector *comprises* a *pol* polynucleotide, which naturally contains the cPPT and CTS regions." Office Action mailed October 10, 2006, at page 4. Applicants disagree with the Office's characterization of Parolin and accordingly, traverse this rejection. Despite the Office's assessment, Parolin does not disclose the complete *pol* sequence and does not disclose the cPPT and CTS sequences required for the claimed invention.

Parolin recites: "The v1864 RSN vector contains gag and pol sequences up to the MseI site (nucleotide 2200) and the v2731 RSN vector contains gag and pol sequences up to the PflmI site (nucleotide 3067)." Parolin at p. 3889. Exhibit A is provided to demonstrate the position of the closest nucleotides recited in Parolin, at page 20, and the beginning of the cPPT and CTS sequences of the claimed invention, at page 26. As shown, nucleotides 2200 and 3067, which are the end-points of the *pol* sequences in the vectors v1864 RSN and v2731 RSN, respectively, disclosed in Parolin are upstream of the cPPT sequence.

Specifically, cPPT is located at the end of the *pol* sequence, inside the integrase sequence, as described in Charneau *et al.* 1991 (Exhibit B) at Figure 1. It should be noted that Charneau 1991 relies on numbering based on the DNA, beginning at the R element from the LTR 5' end, while Parolin relies on numbering based on the RNA beginning at the LTR5' end. This difference results in a shift of 421 base pairs between the numbering. Taking into account this shift, cPPT begins at position 4783, which corresponds to position 4362 in the HXBc2 HIV-1 provirus, the sequence used in Parolin, *see* Parolin at p. 3889, and CTS ends at position 4902, which corresponds to 4481 in HXBc2. Thus, cPPT and CTS begin at 4362, considerably downstream of 3067, the point at which the *pol* sequences in Parolin end. Parolin, then, does not include cPPT or CTS.

Because Parolin does not disclose a transfer vector with the cPPT and CTS sequences, it does not anticipate the claimed invention. Accordingly, Applicants respectfully request that the rejection under 35 U.S.C. § 102(b) in light of Parolin be withdrawn.

Please grant any extensions of time required to enter this response and charge the fee to our Deposit Account No. 06-0916.

Respectfully submitted,

FINNEGAN, HENDERSON, FARABOW,
GARRETT & DUNNER, L.L.P.

Dated: March 9, 2007

By:

Deborah Katz
Reg. No. 51,863
Telephone: (202) 408-4382
Facsimile: (202) 408-4400
E-mail: deborah.katz@finnegan.com

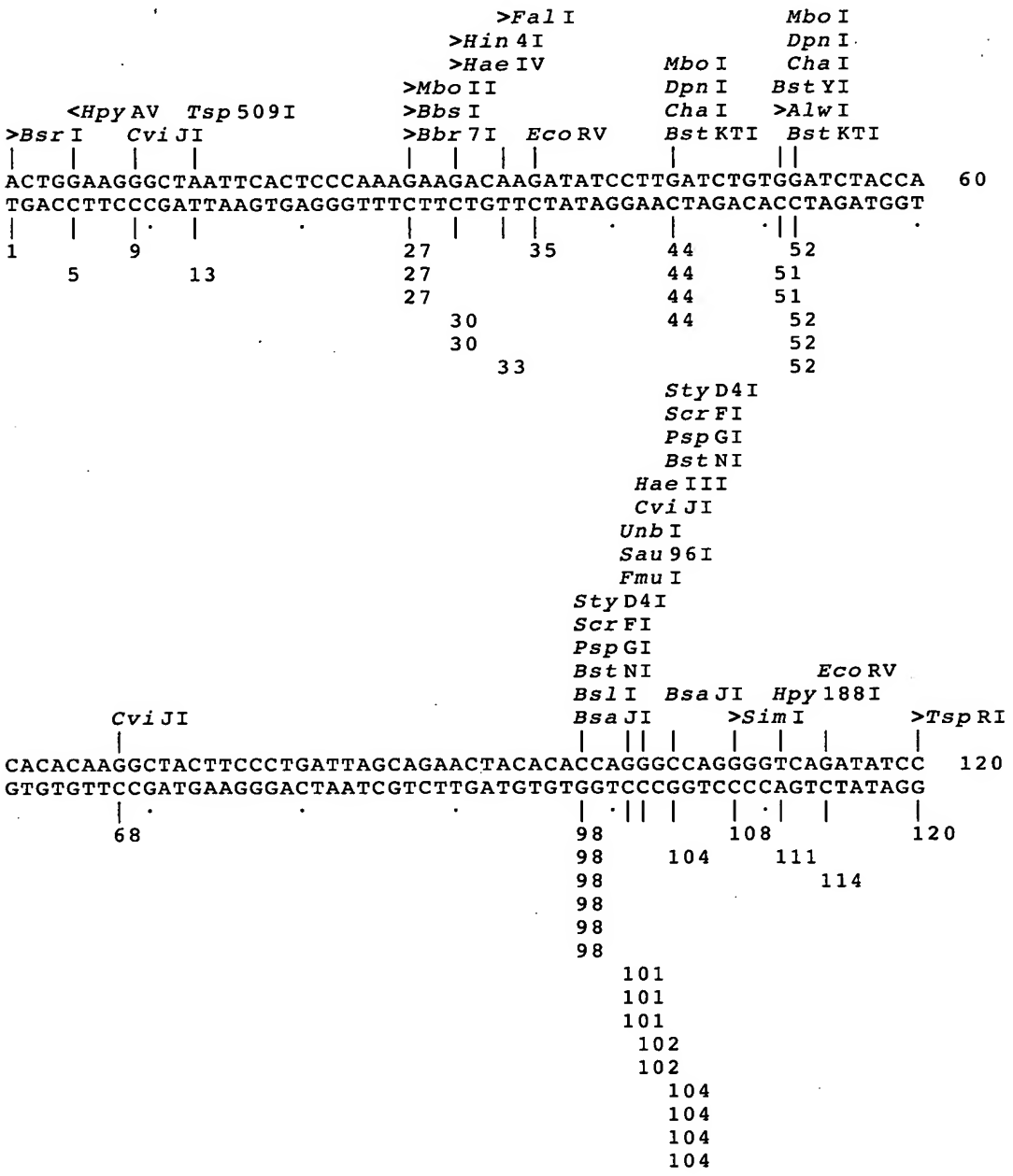
P. Charneau
Feb. 5, 2007

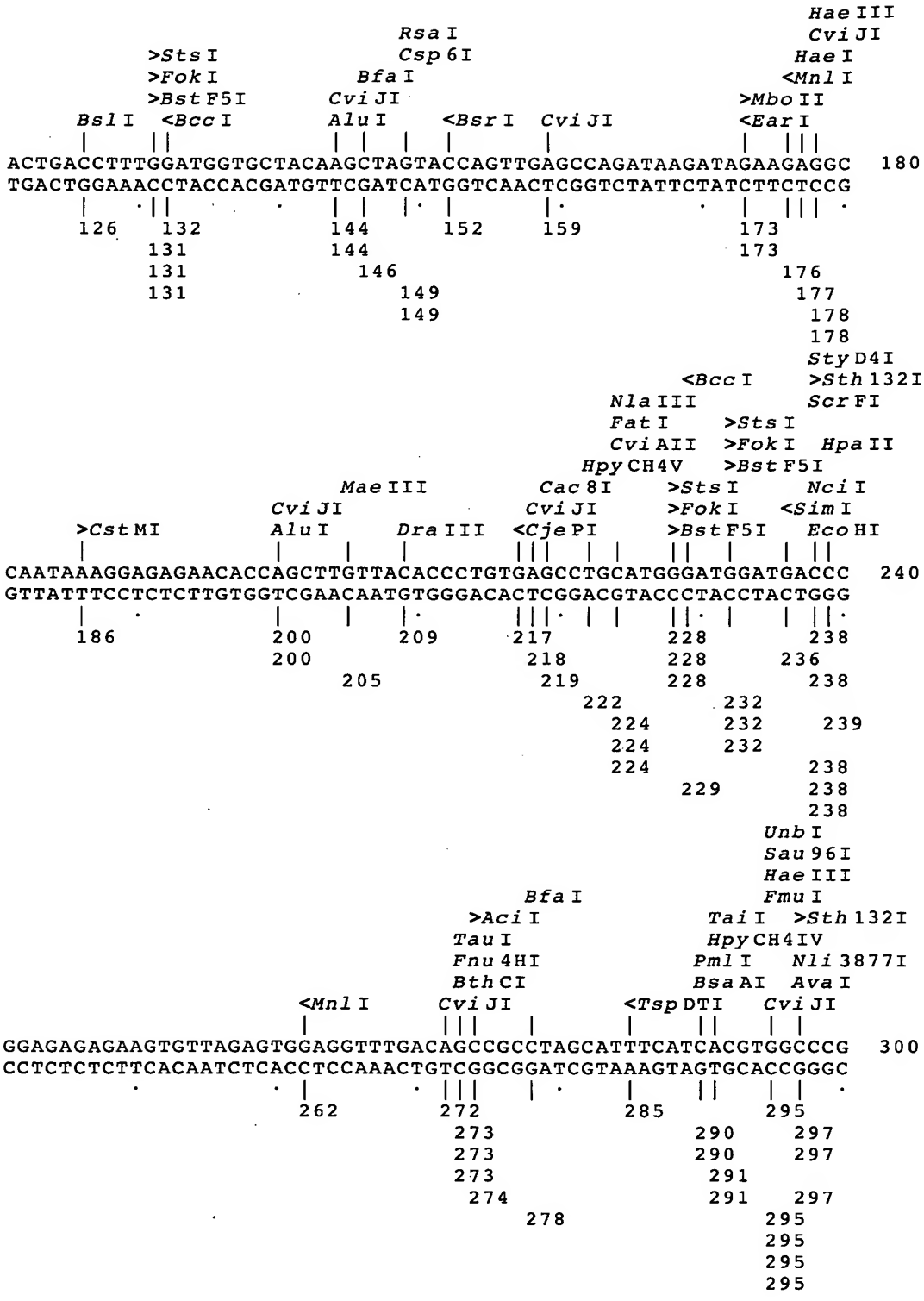
DNA Strider 1.4f1 ### lundi 5 février 2007 16:22:16 (US Letter @ 100%)

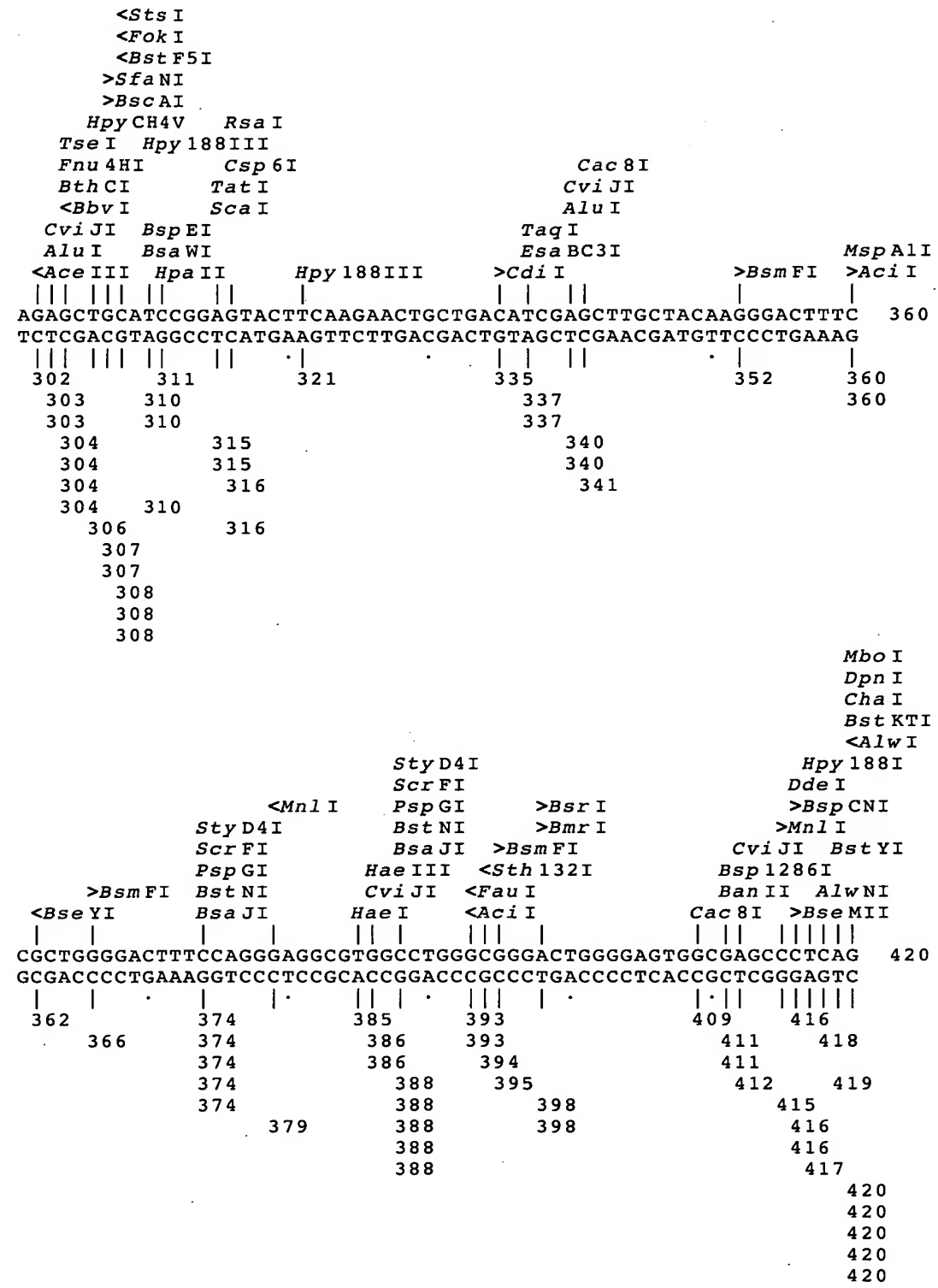
HXB2 -> Restriction Map

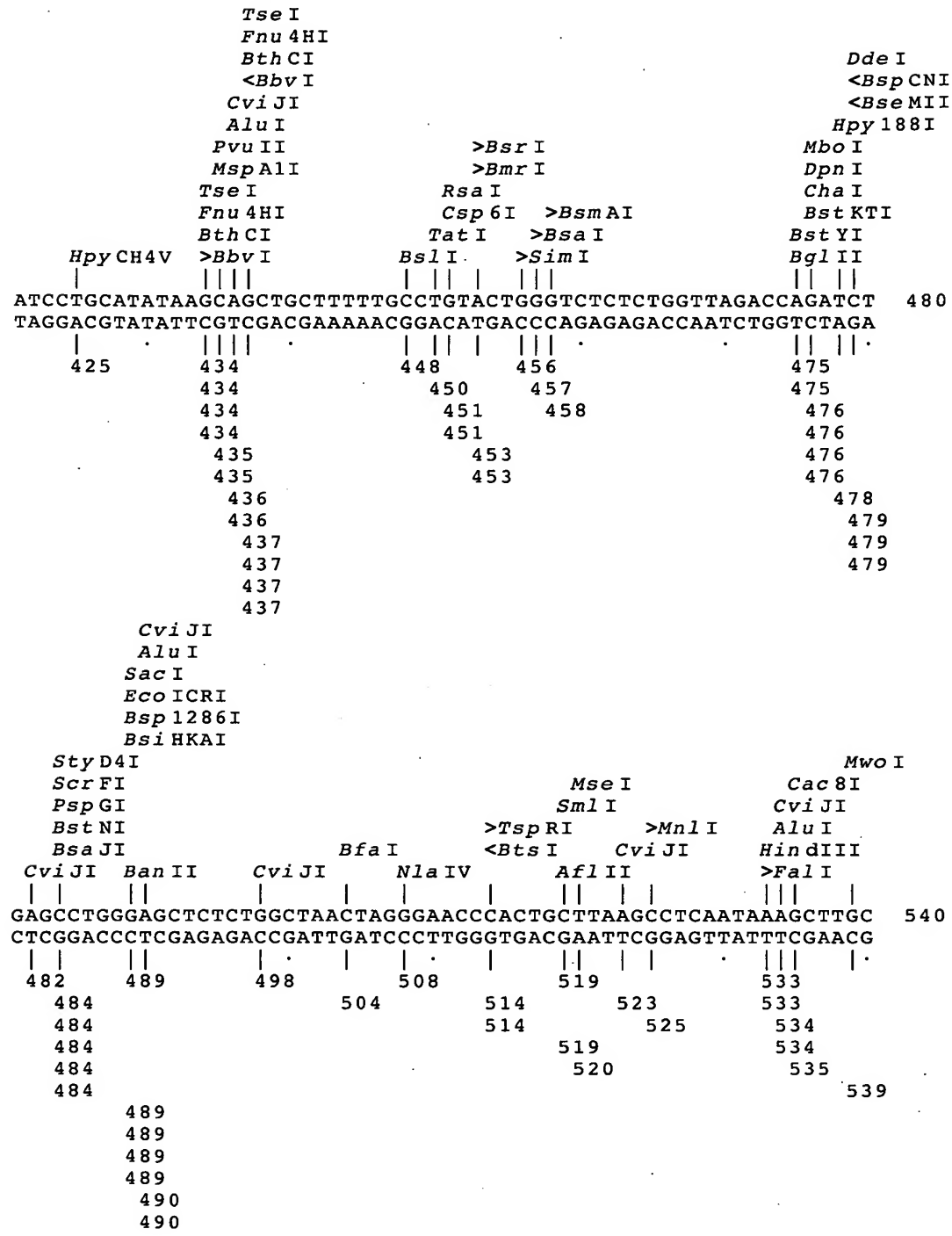
DNA sequence 9721 bp ACTGGAAGGGCT ... AAATCTCTAGCA linear

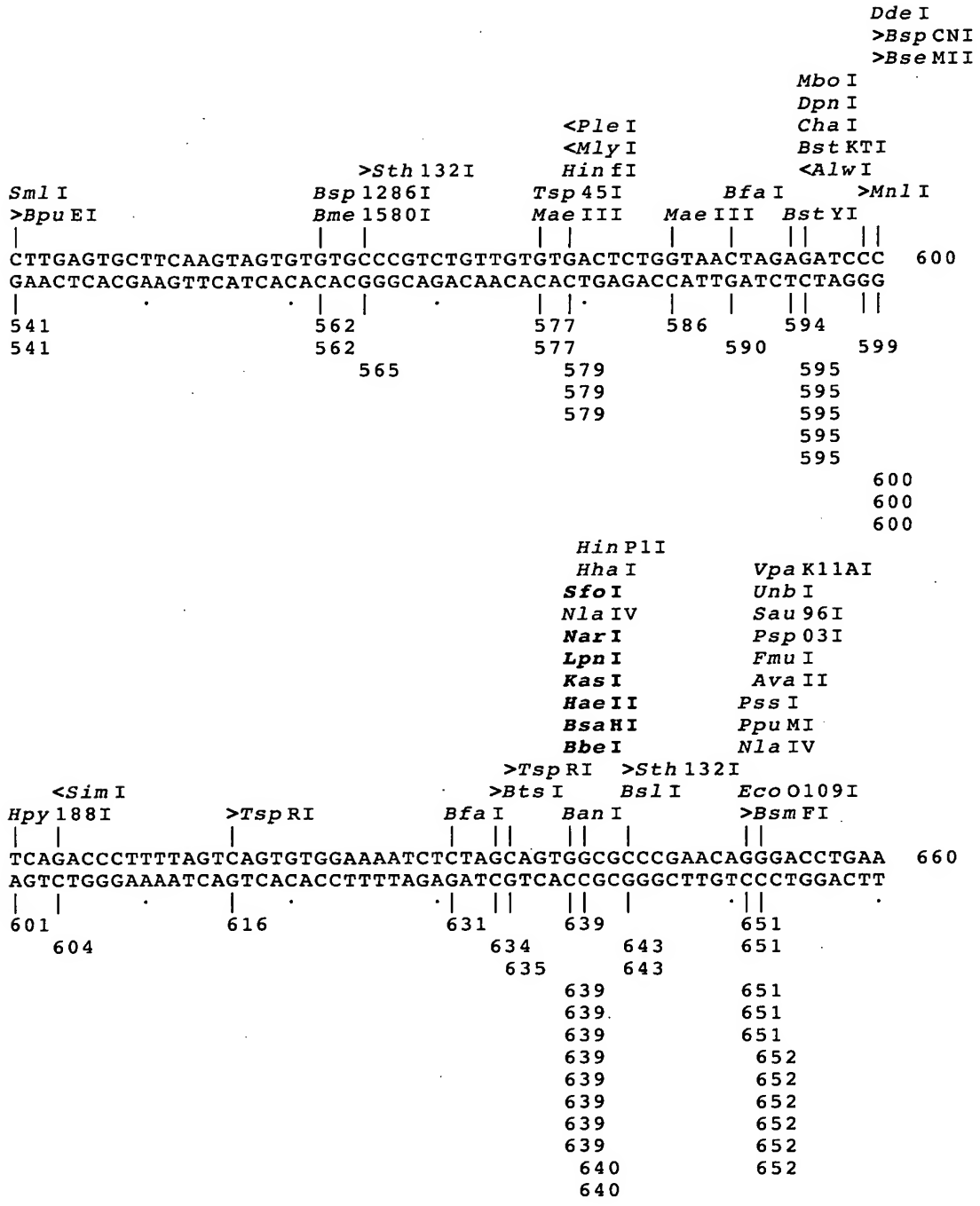
Positions of R. E. sites (sites unique in whole sequence are bold)

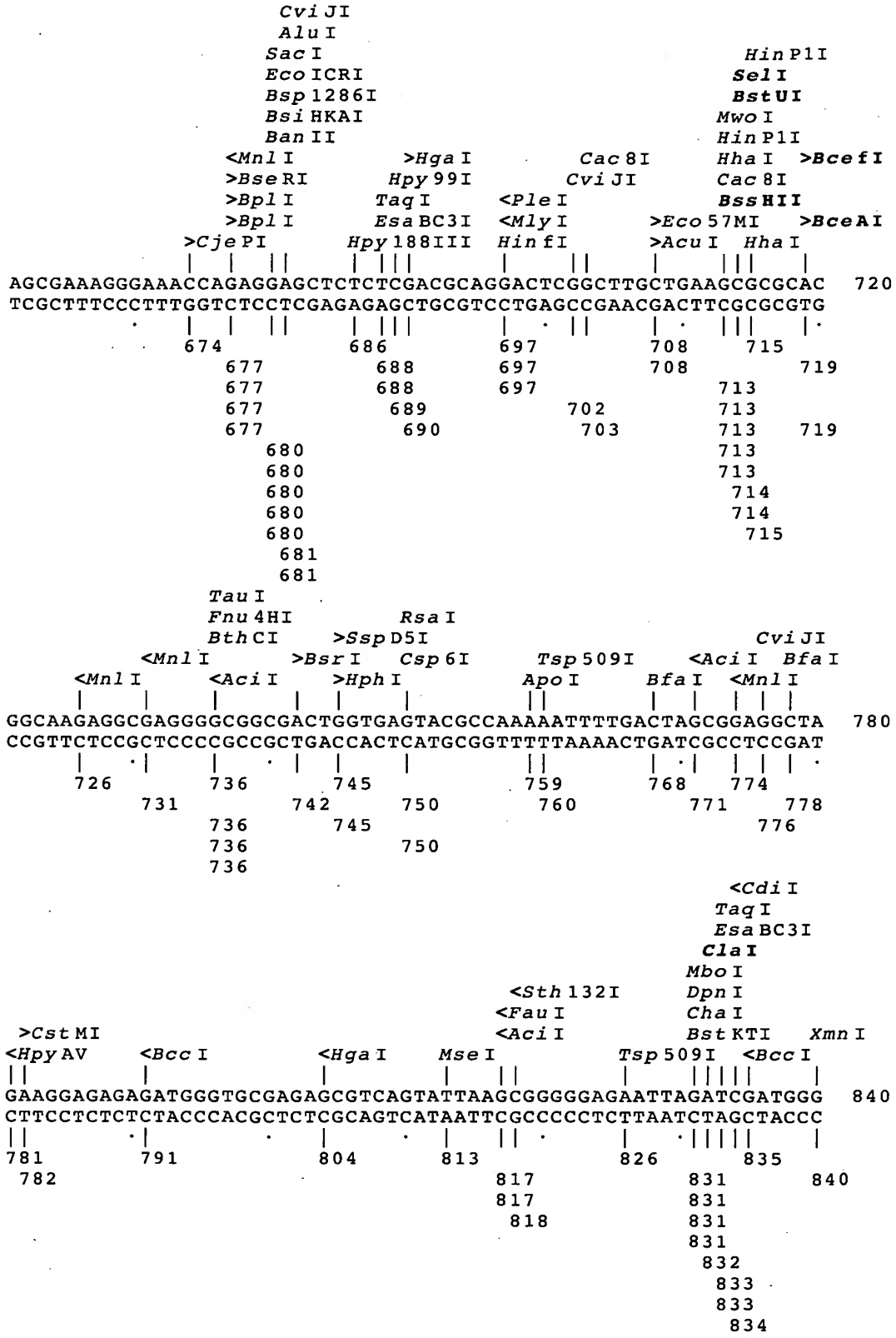


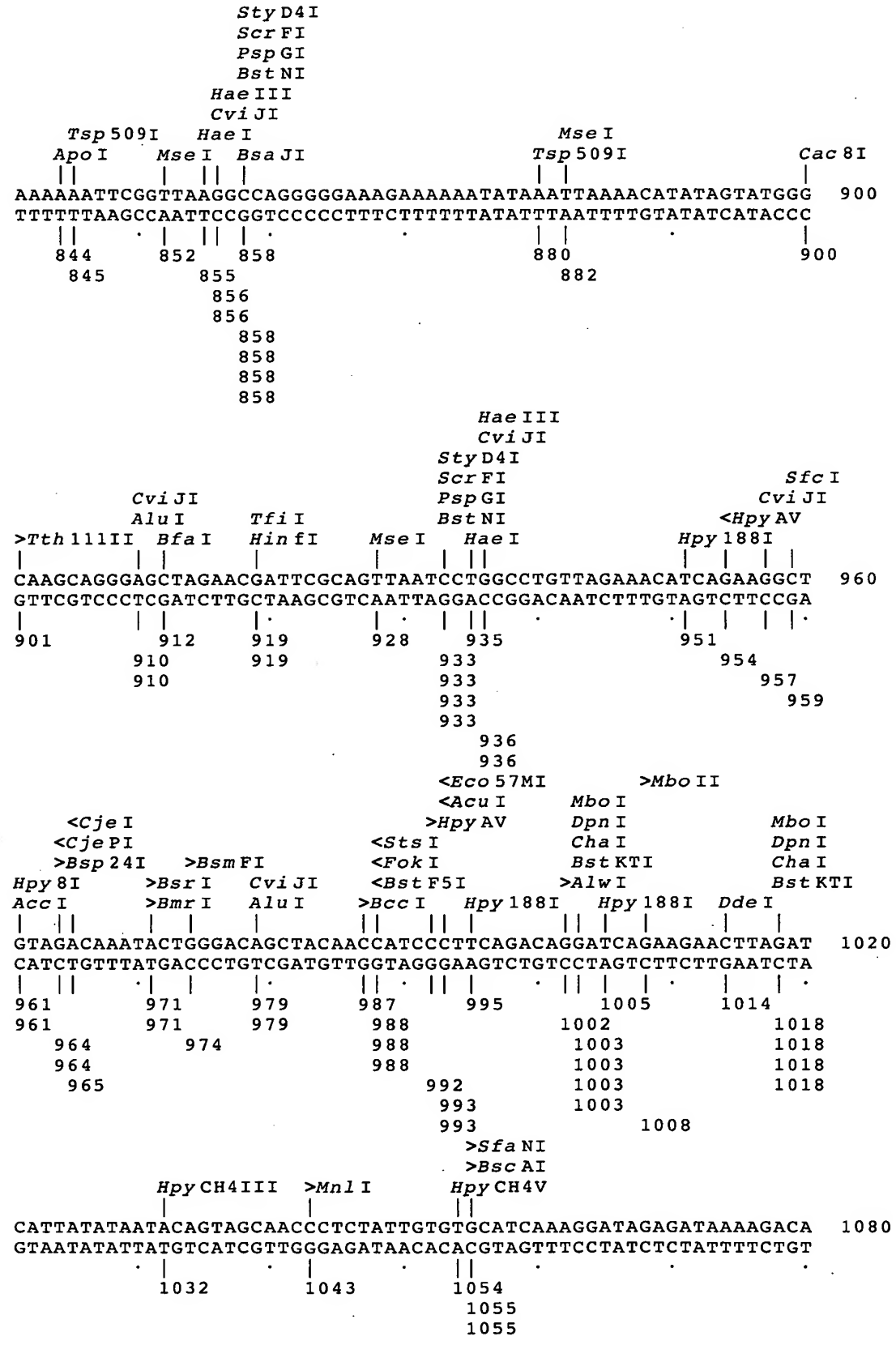

HaeIII
CviJI
RsaI
Csp6I
>StsI
HaeI
>FokI
BfaI
<MnlI
>BstF5I
CviJI
>MboII
BslI
<BccI
AluI
<BsrI
CviJI
<EarI
ACTGACCTTTGGATGGTGCTACAAGCTAGTACCAGTTGAGCCAGATAAGATAGAAGAGGC 180
TGACTGGAAACCTACCACGATGTTCGATCATGGTCAACTCGGTCTATTCTATCTTCTCCG
126 132 131 131 131 144 144 146 149 149 152 159 173 173 176 177 178 178
StyD4I
<BccI
>Sth132I
NlaIII
ScrFI
FatI
>StsI
CviAII
>FokI
HpaII
HpyCH4V
>BstF5I
MaeIII
Cac8I
>StsI
NciI
CviJI
CviJI
>FokI
<SimI
>CstMI
AluI
DraIII
<CjePI
>BstF5I
EcoHI
CAATAAAGGAGAGAACACCAGCTTGTTACACCCTGTGAGCCTGCATGGGATGGATGACCC 240
GTTATTTCCTCTCTTGTGGTCGAACAATGTGGGACACTCGGACGTACCCTACCTACTGGG
186 200 200 205 209 217 218 219 222 224 224 224 228 228 228 229 232 232 232 236 238 238 238 238 239
UnbI
Sau96I
HaeIII
BfaI
FmuI
>AciI
TaiI
>Sth132I
TauI
HpyCH4IV
Fnu4HI
PmlI
Nli3877I
BthCI
BsaAI
AvaI
<MnlI
CviJI
<TspDTI
CviJI
GGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCATTTCATCACGTGGCCCG 300
CCTCTCTCTTCACAATCTCACCTCCAAACTGTCGGCGGATCGTAAAGTAGTGCACCGGGC
262 272 273 273 273 274 278 285 290 290 291 291 295 295 295 295 295 297 297 297

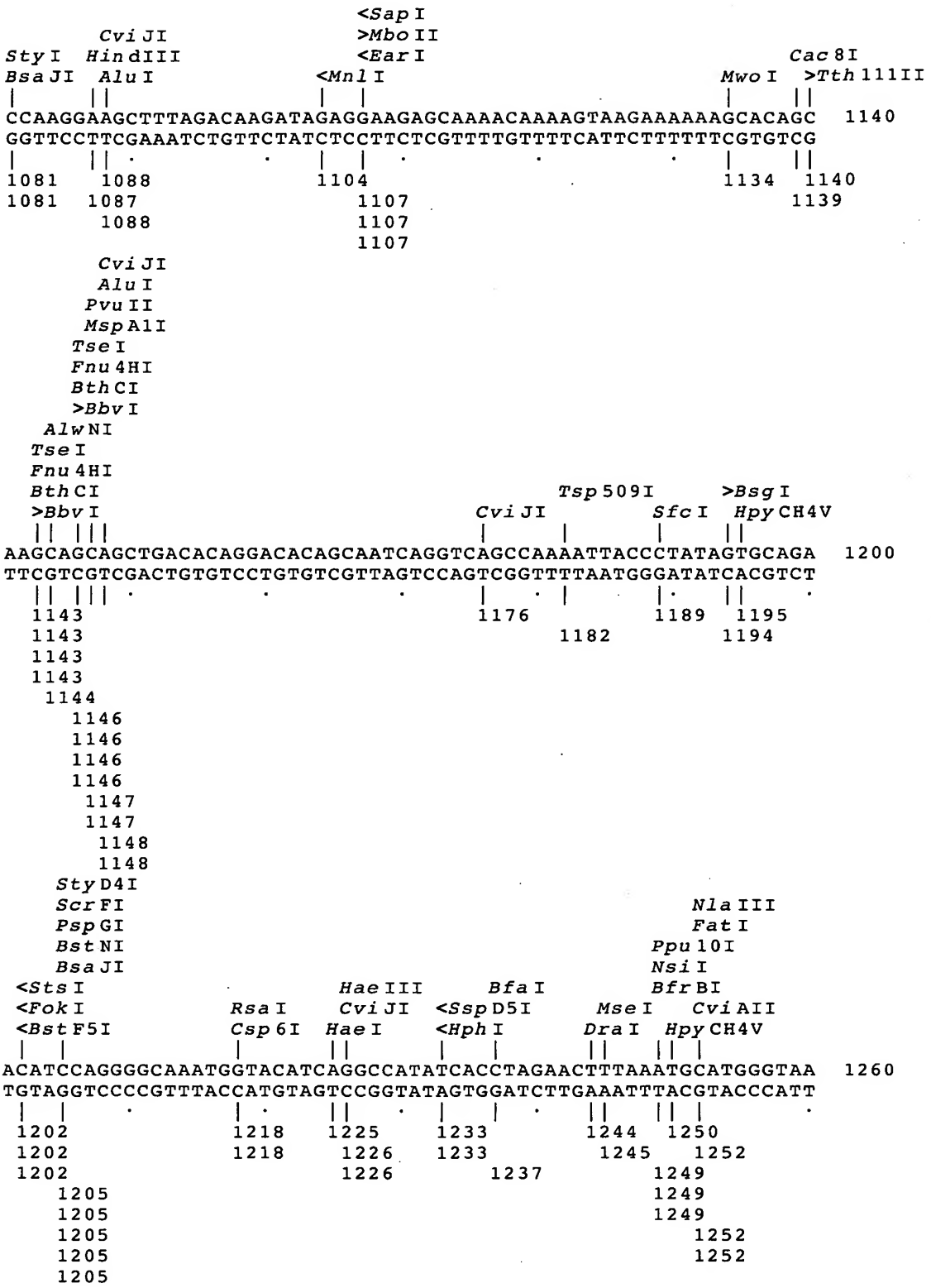

<Sts I
<Fok I
<Bst F5I
>Sfa NI
>Bsc AI
Hpy CH4V Rsa I
Tse I Hpy 188III
Fnu 4HI Csp 6I Cac 8I
Bth CI Tat I Cvi JI
<Bbv I Sca I Alu I
Cvi JI Bsp EI Taq I
Alu I Bsa WI Esa BC3I Msp A1I
<Ace III Hpa II Hpy 188III >Cdi I >Bsm FI >Aci I
AGAGCTGCATCCGGAGTACTTCAAGAACTGCTGACATCGAGCTTGCTACAAGGGACTTTC 360
TCTCGACGTAGGCCTCATGAAGTTCTTGACGACTGTAGCTCGAACGATGTTCCCTGAAAG
302 311 321 335 352 360
303 310 337 360
303 310 337
304 315 340
304 315 340
304 316 341
304 310
306 316
307
307
308
308
308
Mbo I
Dpn I
Cha I
Bst KTI
<Alw I
Sty D4I Hpy 188I
Scr FI Dde I
<Mnl I Psp GI >Bsr I >Bsp CNI
Sty D4I Bst NI >Bmr I >Mnl I
Scr FI Bsa JI >Bsm FI Cvi JI Bst YI
Psp GI Hae III <Sth 132I Bsp 1286I
>Bsm FI Bst NI Cvi JI <Fau I Ban II Alw NI
<Bse YI Bsa JI Hae I <Aci I Cac 8I >Bse MII
CGCTGGGGACTTTCCAGGGAGGCGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAG 420
GCGACCCCTGAAAGGTCCCTCCGCACCGGACCCGCCCTGACCCCTCACCGCTCGGGAGTC
362 374 385 393 409 416
366 374 386 393 411 418
374 386 394 411
374 388 395 412 419
374 388 398 415
379 388 398 416
388 416
388 417
420
420
420
420
420

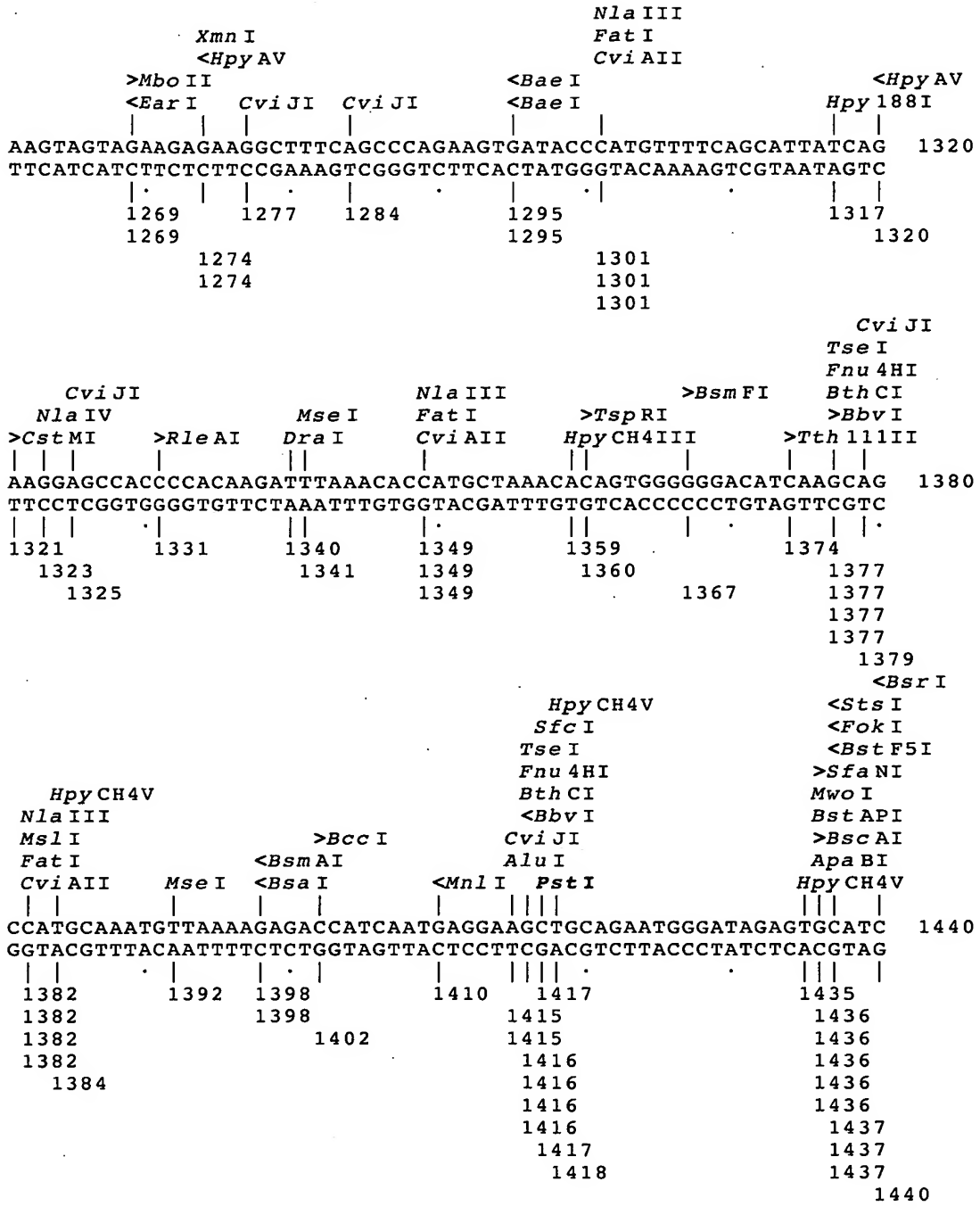

Tse I
Fnu 4HI
Bth CI
<Bbv I
Cvi JI
Alu I
Pvu II
Msp A1I
Tse I
Fnu 4HI
Bth CI
Hpy CH4V
>Bbv I
>Bsr I
>Bmr I
Rsa I
Csp 6I
>Bsm AI
Tat I
>Bsa I
Bsl I
>Sim I
Dde I
<Bsp CNI
<Bse MII
Hpy 188I
Mbo I
Dpn I
Cha I
Bst KTI
Bst YI
Bgl II
ATCCTGCATATAAGCAGCTGCTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATCT 480
TAGGACGTATATTCGTCGACGAAAAACGGACATGACCCAGAGAGACCAATCTGGTCTAGA
425
434
434
434
434
435
435
436
436
437
437
437
437
448
450
451
451
453
453
456
457
458
475
475
476
476
476
476
478
479
479
479
Cvi JI
Alu I
Sac I
Eco ICRI
Bsp 1286I
Bsi HKAI
Sty D4I
Scr FI
Psp GI
Bst NI
Bsa JI
Bfa I
Mse I
Sml I
>Tsp RI
<Bts I
>Mnl I
Cvi JI
Cac 8I
Cvi JI
Alu I
Hin dIII
>Fal I
Mwo I
Cvi JI
Ban II
Cvi JI
Nla IV
Afl II
GAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGC 540
CTCGGACCCTCGAGAGACCGATTGATCCCTTGGGTGACGAATTCGGAGTTATTTCGAACG
482
484
484
484
484
484
489
489
489
489
489
490
490
498
504
508
514
514
519
520
523
525
533
533
534
534
535
539

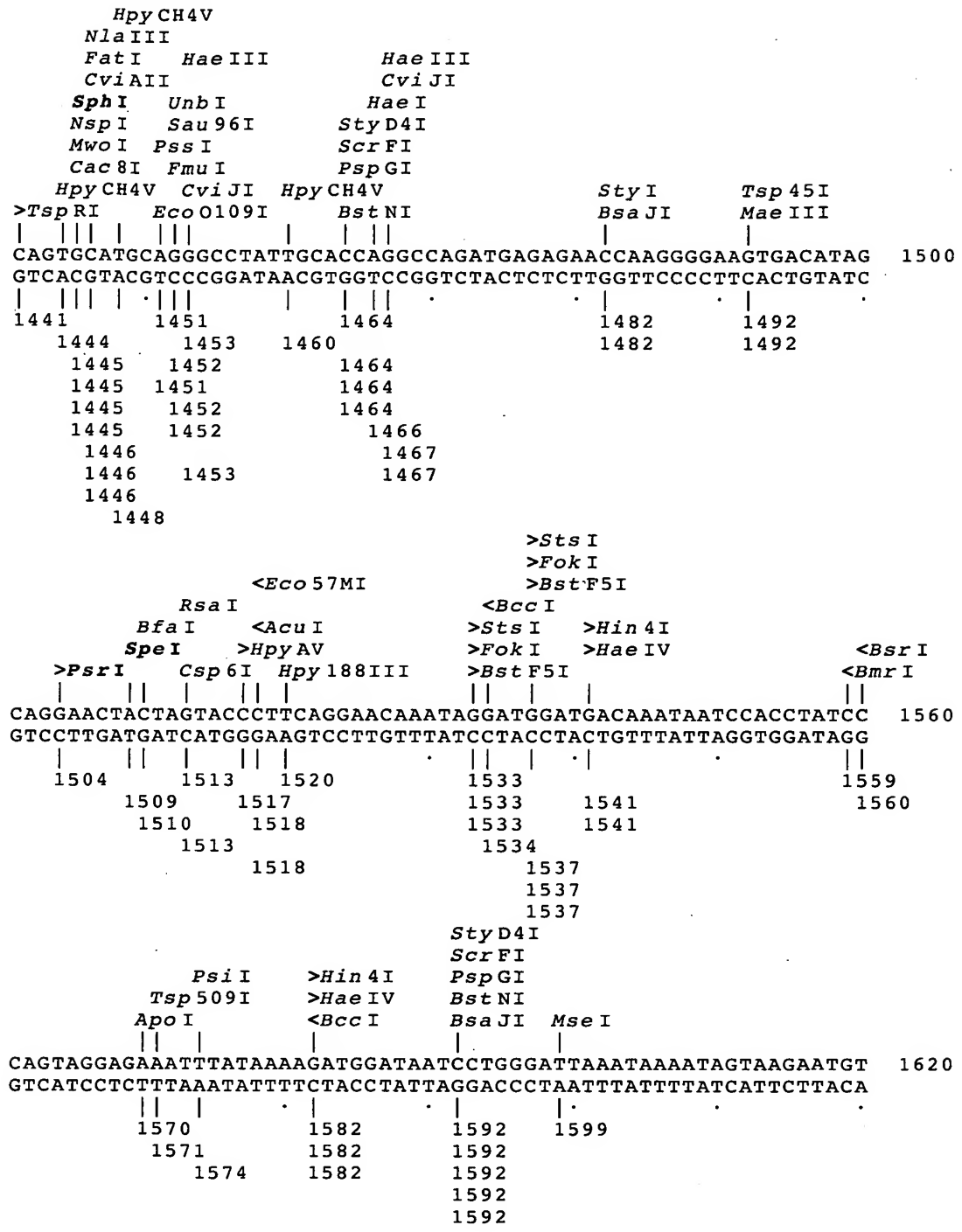

Dde I
>Bsp CNI
>Bse MII
Mbo I
Dpn I
Cha I
Bst KTI
<Alw I
<Ple I
<Mly I
Hin fI
>Sth 132I
Sml I
>Bpu EI
Bsp 1286I
Bme 1580I
Tsp 45I
Mae III
Mae III
Bfa I
Bst YI
>Mnl I
CTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCC 600
GAACTCACGAAGTTCATCACACACGGGCAGACAACACACTGAGACCATTGATCTCTAGGG
541 562 577 586 594
541 562 577 590 599
565 579 595
579 595
579 595
595
595
600
600
600
Hin P1I
Hha I
Sfo I
Nla IV
Nar I
Lpn I
Kas I
Hae II
Bsa HI
Bbe I
Vpa K11AI
Unb I
Sau 96I
Psp 03I
Fmu I
Ava II
Pss I
Ppu MI
Nla IV
>Tsp RI
>Sth 132I
<Sim I
>Bts I
Bsl I
Eco O109I
Hpy 188I
>Tsp RI
Bfa I
Ban I
>Bsm FI
TCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGCGCCCGAACAGGGACCTGAA 660
AGTCTGGGAAAATCAGTCACACCTTTTAGAGATCGTCACCGCGGGCTTGTCCCTGGACTT
601 616 631 639 651
604 634 643 651
635 643
639 651
639 651
639 651
639 652
639 652
639 652
639 652
639 652
640 652
640

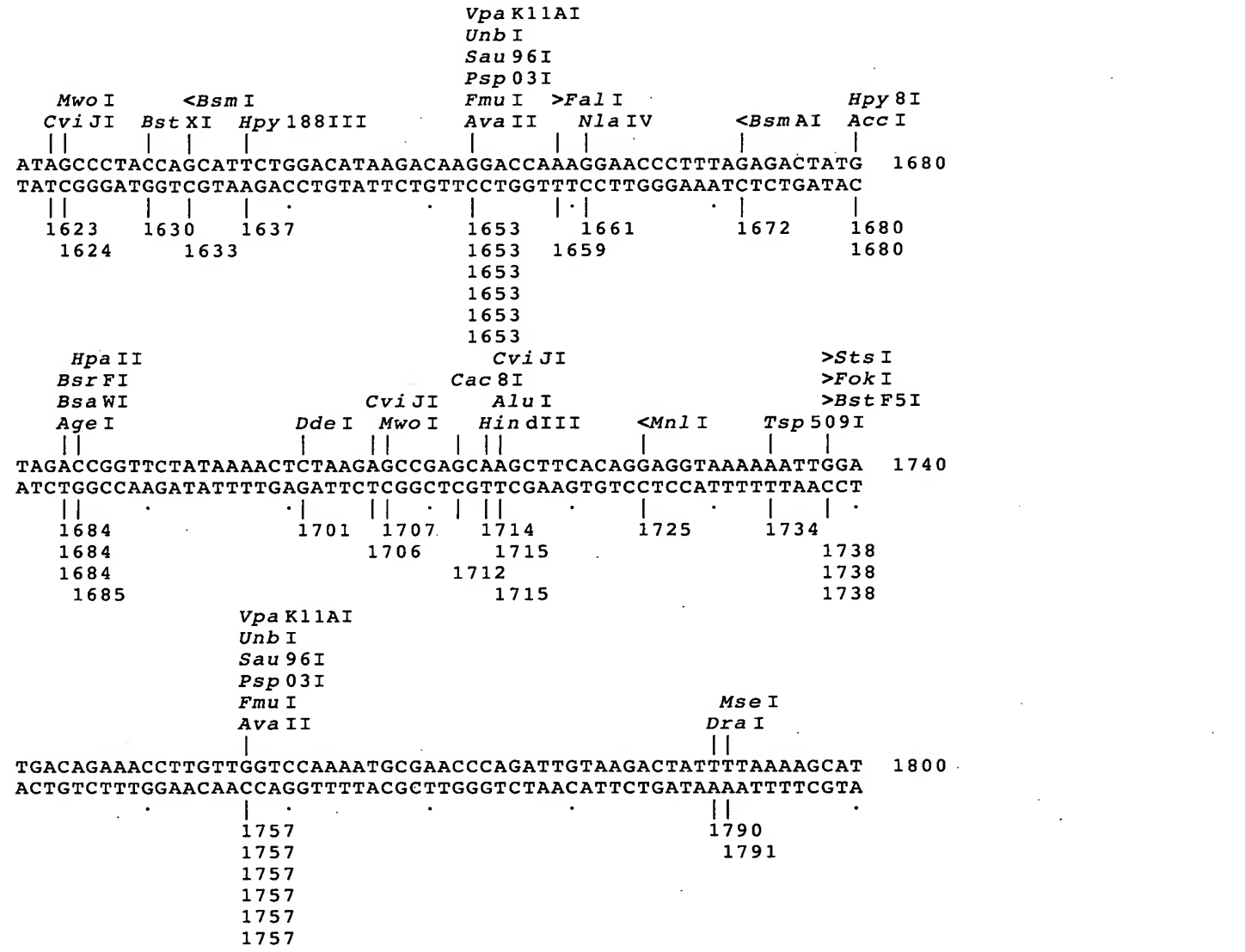

CviJI
AluI
SacI
EcoICRI
Bsp1286I
BsiHKAI
BanII
HinP1I
SelI
BstUI
MwoI
HinP1I
<MnlI
>BseRI
>BplI
>BplI
>HgaI
Hpy99I
TaqI
EsaBC3I
Cac8I
CviJI
HhaI
>BcefI
Cac8I
BssHII
<PleI
<MlyI
HinfI
>Eco57MI
>BceAI
>AcuI
HhaI
>CjePI
Hpy188III
AGCGAAAGGGAAACCAGAGGAGCTCTCTCGACGCAGGACTCGGCTTGCTGAAGCGCGCAC 720
TCGCTTTCCCTTTGGTCTCCTCGAGAGAGCTGCGTCCTGAGCCGAACGACTTCGCGCGTG
674 686 697 697 708 708 715
677 688 697 708 719
677 688 697 713
677 689 702 713
677 690 703 713 719
680 713
680 713
680 714
680 714
680 715
681
681
TauI
Fnu4HI
BthCI
RsaI
>SspD5I
CviJI
<MnlI
>BsrI
CspI
Tsp509I
<AciI
BfaI
<MnlI
<MnlI
<AciI
>HphI
ApoI
BfaI
<MnlI
GGCAAGAGGCGAGGGGCGGCGACTGGTGAGTACGCCAAAAATTTTGACTAGCGGAGGCTA 780
CCGTTCTCCGCTCCCCGCCGCTGACCACTCATGCGGTTTTTAAAACTGATCGCCTCCGAT
726 731 736 736 736 742 745 745 750 750 759 760 768 771 774 776 778
<CdiI
TaqI
EsaBC3I
ClaI
MboI
DpnI
ChaI
BstKTI
XmnI
<Sth132I
<FauI
<AciI
>CstMI
<HpyAV
<BccI
<HgaI
MseI
Tsp509I
<BccI
GAAGGAGAGAGATGGGTGCGAGAGCGTCAGTATTAAGCGGGGGAGAATTAGATCGATGGG 840
CTTCCTCTCTCTACCCACGCTCTCGCAGTCATAATTCGCCCCCTCTTAATCTAGCTACCC
781 782 791 804 813 817 817 818 826 831 831 831 831 832 833 833 834 835 840

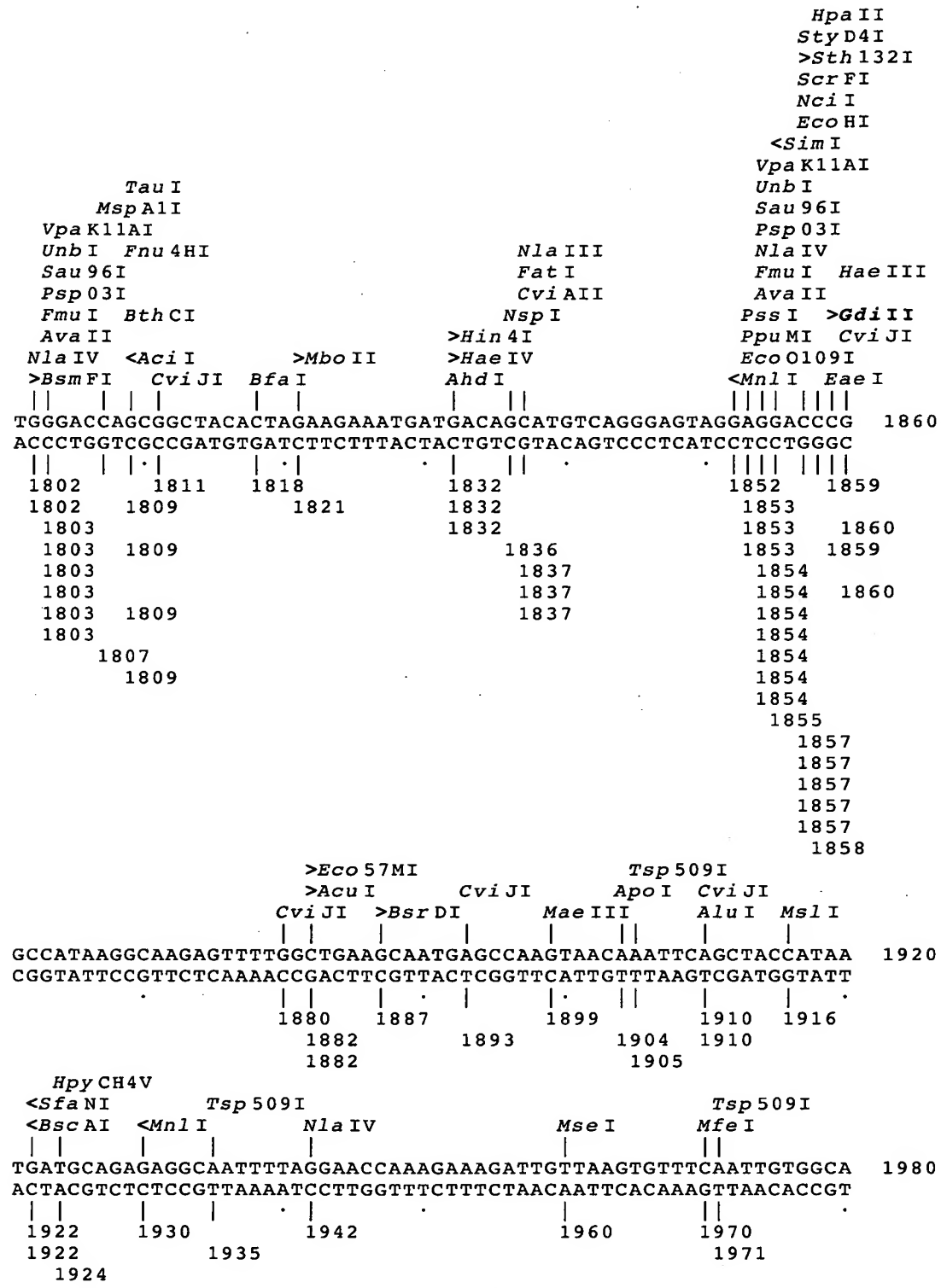

Sty D4I
Scr FI
Psp GI
Bst NI
Hae III
Cvi JI
Tsp 509I
Hae I
Mse I
Apo I
Mse I
Bsa JI
Tsp 509I
Cac 8I
AAAAAATTCGGTTAAGGCCAGGGGGAAAGAAAAAATATAAATTAAAACATATAGTATGGG 900
TTTTTTAAGCCAATTCCGGTCCCCCTTTCTTTTTTATATTTAATTTTGTATATCATACCC
844 852 858 880 900
845 855 882
856
856
858
858
858
858
Hae III
Cvi JI
Sty D4I
Scr FI
Psp GI
Bst NI
Sfc I
Cvi JI
Cvi JI
Alu I
Tfi I
<Hpy AV
>Tth 111II
Bfa I
Hin fI
Mse I
Hae I
Hpy 188I
CAAGCAGGGAGCTAGAACGATTCGCAGTTAATCCTGGCCTGTTAGAAACATCAGAAGGCT 960
GTTCGTCCCTCGATCTTGCTAAGCGTCAATTAGGACCGGACAATCTTTGTAGTCTTCCGA
901 912 919 928 935 951
910 919 933 954
910 933 957
933 959
933
936
936
<Eco 57MI
>Mbo II
<Cje I
<Acu I
Mbo I
<Cje PI
>Hpy AV
Dpn I
Mbo I
>Bsp 24I
>Bsm FI
<Sts I
<Fok I
Cha I
Dpn I
Hpy 8I
>Bsr I
Cvi JI
<Bst F5I
Bst KTI
>Alw I
Cha I
Acc I
>Bmr I
Alu I
>Bcc I
Hpy 188I
Hpy 188I
Dde I
Bst KTI
GTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAAGAACTTAGAT 1020
CATCTGTTTATGACCCTGTCGATGTTGGTAGGGAAGTCTGTCCTAGTCTTCTTGAATCTA
961 971 979 987 995 1005 1014
961 971 979 988 1002 1018
964 974 988 1003 1018
964 988 1003 1018
965 992 1003 1018
993 1003
993 1008
>Sfa NI
>Bsc AI
Hpy CH4III
>Mnl I
Hpy CH4V
CATTATATAATACAGTAGCAACCCTCTATTGTGTGCATCAAAGGATAGAGATAAAAGACA 1080
GTAATATATTATGTCATCGTTGGGAGATAACACACGTAGTTTCCTATCTCTATTTTCTGT
1032 1043 1054
1055
1055

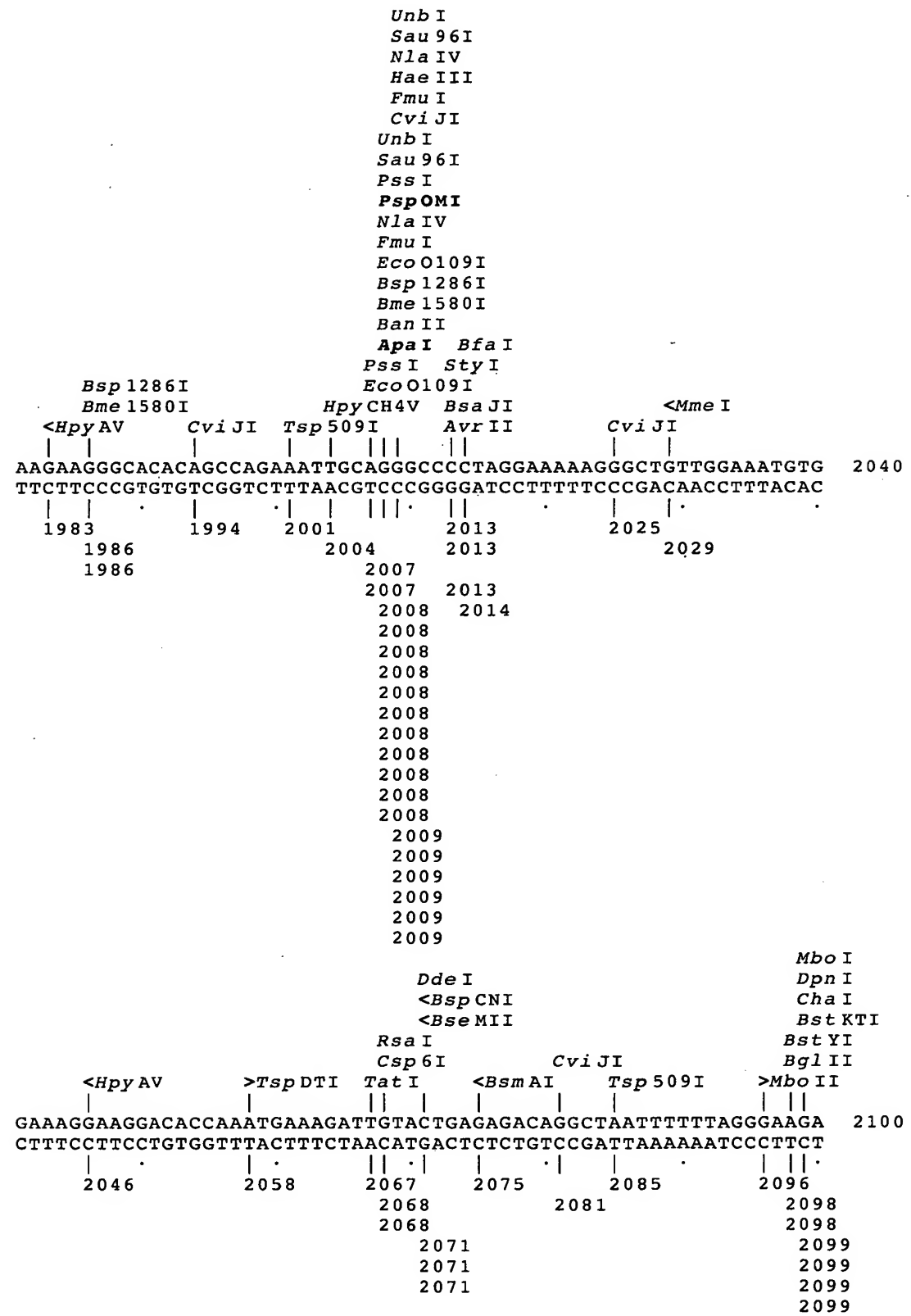

<Sap I
>Mbo II
<Ear I
Cvi JI
Sty I Hin dIII
Bsa JI Alu I
<Mnl I
Mwo I
Cac 8I
>Tth 111II
CCAAGGAAGCTTTAGACAAGATAGAGGAAGAGCAAAACAAAAGTAAGAAAAAAGCACAGC 1140
GGTTCCTTCGAAATCTGTTCTATCTCCTTCTCGTTTTGTTTTCATTCTTTTTTCGTGTCG
1081 1088 1104 1134 1140
1081 1087 1107 1139
1088 1107
1107
Cvi JI
Alu I
Pvu II
Msp A1I
Tse I
Fnu 4HI
Bth CI
>Bbv I
Alw NI
Tse I
Fnu 4HI
Bth CI
>Bbv I
Tsp 509I
>Bsg I
Cvi JI
Sfc I
Hpy CH4V
AAGCAGCAGCTGACACAGGACACAGCAATCAGGTCAGCCAAAATTACCCTATAGTGCAGA 1200
TTCGTCGTCGACTGTGTCCTGTGTCGTTAGTCCAGTCGGTTTTAATGGGATATCACGTCT
1143 1176 1189 1195
1143 1182 1194
1143
1143
1144
1146
1146
1146
1146
1147
1147
1148
1148
Sty D4I
Scr FI
Psp GI
Bst NI
Bsa JI
Nla III
Fat I
Ppu 10I
Nsi I
<Sts I
<Fok I
<Bst F5I
Rsa I
Csp 6I
Hae III
Cvi JI
Hae I
<Ssp D5I
<Hph I
Bfa I
Mse I
Dra I
Bfr BI
Cvi AII
Hpy CH4V
ACATCCAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAATGCATGGGTAA 1260
TGTAGGTCCCCGTTTACCATGTAGTCCGGTATAGTGGATCTTGAAATTTACGTACCCATT
1202 1218 1225 1233 1244 1250
1202 1218 1226 1233 1245 1252
1202 1226 1237 1249
1205 1249
1205 1249
1205 1252
1205 1252
1205

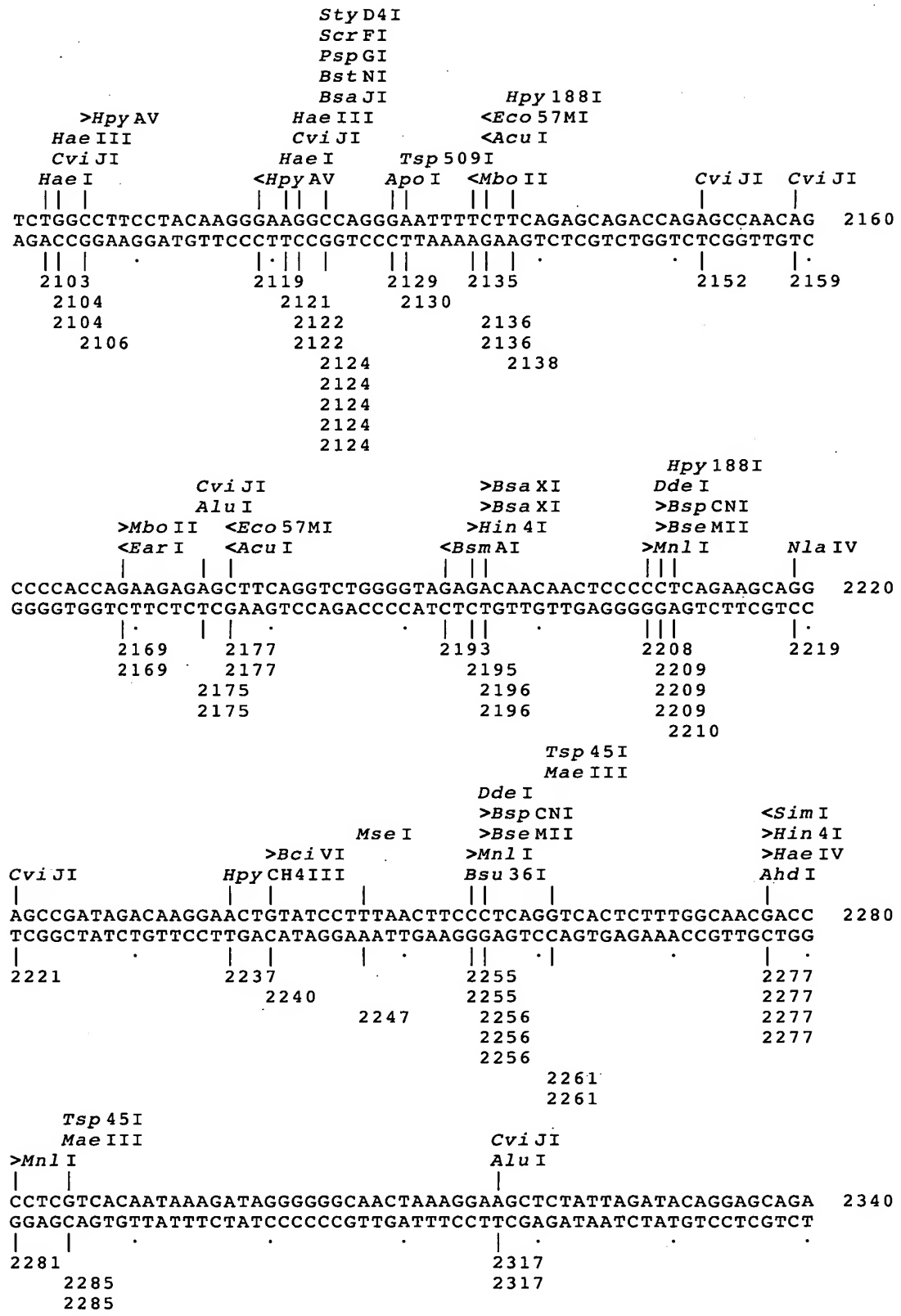

Nla III
Fat I
Cvi AII
Xmn I
<Hpy AV
>Mbo II
<Ear I
Cvi JI
Cvi JI
<Bae I
<Bae I
<Hpy AV
Hpy 188I
AAGTAGTAGAAGAGAAGGCTTTCAGCCCAGAAGTGATACCCATGTTTTCAGCATTATCAG 1320
TTCATCATCTTCTCTTCCGAAAGTCGGGTCTTCACTATGGGTACAAAAGTCGTAATAGTC
1269
1269
1277
1284
1295
1295
1317
1320
1274
1274
1301
1301
1301
Cvi JI
Tse I
Fnu 4HI
Cvi JI
Nla III
>Bsm FI
Bth CI
Nla IV
Mse I
Fat I
>Tsp RI
>Bbv I
>Cst MI
>Rle AI
Dra I
Cvi AII
Hpy CH4III
>Tth 111II
AAGGAGCCACCCCACAAGATTTAAACACCATGCTAAACACAGTGGGGGGACATCAAGCAG 1380
TTCCTCGGTGGGGTGTTCTAAATTTGTGGTACGATTTGTGTCACCCCCCTGTAGTTCGTC
1321
1331
1340
1349
1359
1374
1323
1341
1349
1360
1377
1325
1349
1367
1377
1377
1377
1379
<Bsr I
Hpy CH4V
<Sts I
Sfc I
<Fok I
Tse I
<Bst F5I
Fnu 4HI
>Sfa NI
Hpy CH4V
Bth CI
Mwo I
Nla III
<Bbv I
Bst API
Msl I
>Bcc I
Cvi JI
>Bsc AI
Fat I
<Bsm AI
Alu I
Apa BI
Cvi AII
Mse I
<Bsa I
<Mnl I
Pst I
Hpy CH4V
CCATGCAAATGTTAAAAGAGACCATCAATGAGGAAGCTGCAGAATGGGATAGAGTGCATC 1440
GGTACGTTTACAATTTTCTCTGGTAGTTACTCCTTCGACGTCTTACCCTATCTCACGTAG
1382
1392
1398
1410
1417
1435
1382
1398
1415
1436
1382
1402
1415
1436
1382
1416
1436
1384
1416
1436
1416
1436
1416
1437
1417
1437
1418
1437
1440

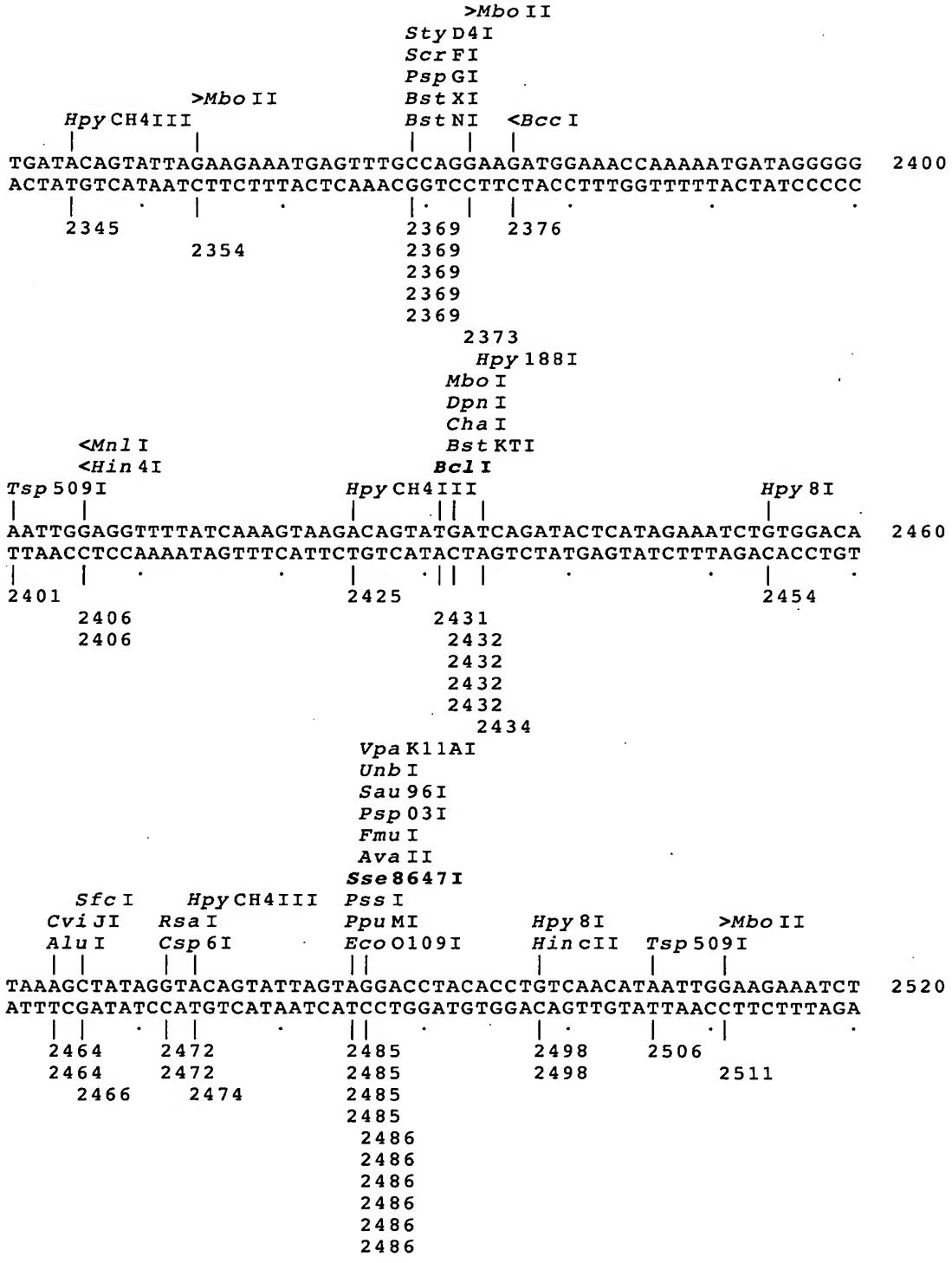

HpyCH4V
NlaIII
FatI HaeIII HaeIII
CviAII CviJI
SphI UnbI HaeI
NspI Sau96I StyD4I
MwoI PssI ScrFI
Cac8I FmuI PspGI
HpyCH4V CviJI HpyCH4V StyI Tsp45I
>TspRI Eco0109I BstNI BsaJI MaeIII
CAGTGCATGCAGGGCCTATTGCACCAGGCCAGATGAGAGAACCAAGGGGAAGTGACATAG 1500
GTCACGTACGTCCCGGATAACGTGGTCCGGTCTACTCTCTTGGTTCCCCTTCACTGTATC
1441 1451 1464 1482 1492
1444 1453 1460 1482 1492
1445 1452 1464
1445 1451 1464
1445 1452 1464
1445 1452 1466
1446 1467
1446 1453 1467
1446
1448
>StsI
>FokI
<Eco57MI >BstF5I
RsaI <BccI
BfaI <AcuI >StsI >Hin4I
SpeI >HpyAV >FokI >HaeIV <BsrI
>PsrI Csp6I Hpy188III >BstF5I <BmrI
CAGGAACTACTAGTACCCTTCAGGAACAAATAGGATGGATGACAAATAATCCACCTATCC 1560
GTCCTTGATGATCATGGGAAGTCCTTGTTTATCCTACCTACTGTTTATTAGGTGGATAGG
1504 1513 1520 1533 1559
1509 1517 1533 1541 1560
1510 1518 1533 1541
1513 1534
1518 1537
1537
1537
StyD4I
ScrFI
PsiI >Hin4I PspGI
Tsp509I >HaeIV BstNI
ApoI <BccI BsaJI MseI
CAGTAGGAGAAATTTATAAAAGATGGATAATCCTGGGATTAAATAAAATAGTAAGAATGT 1620
GTCATCCTCTTTAAATATTTTCTACCTATTAGGACCCTAATTTATTTTATCATTCTTACA
1570 1582 1592 1599
1571 1582 1592
1574 1582 1592
1592
1592

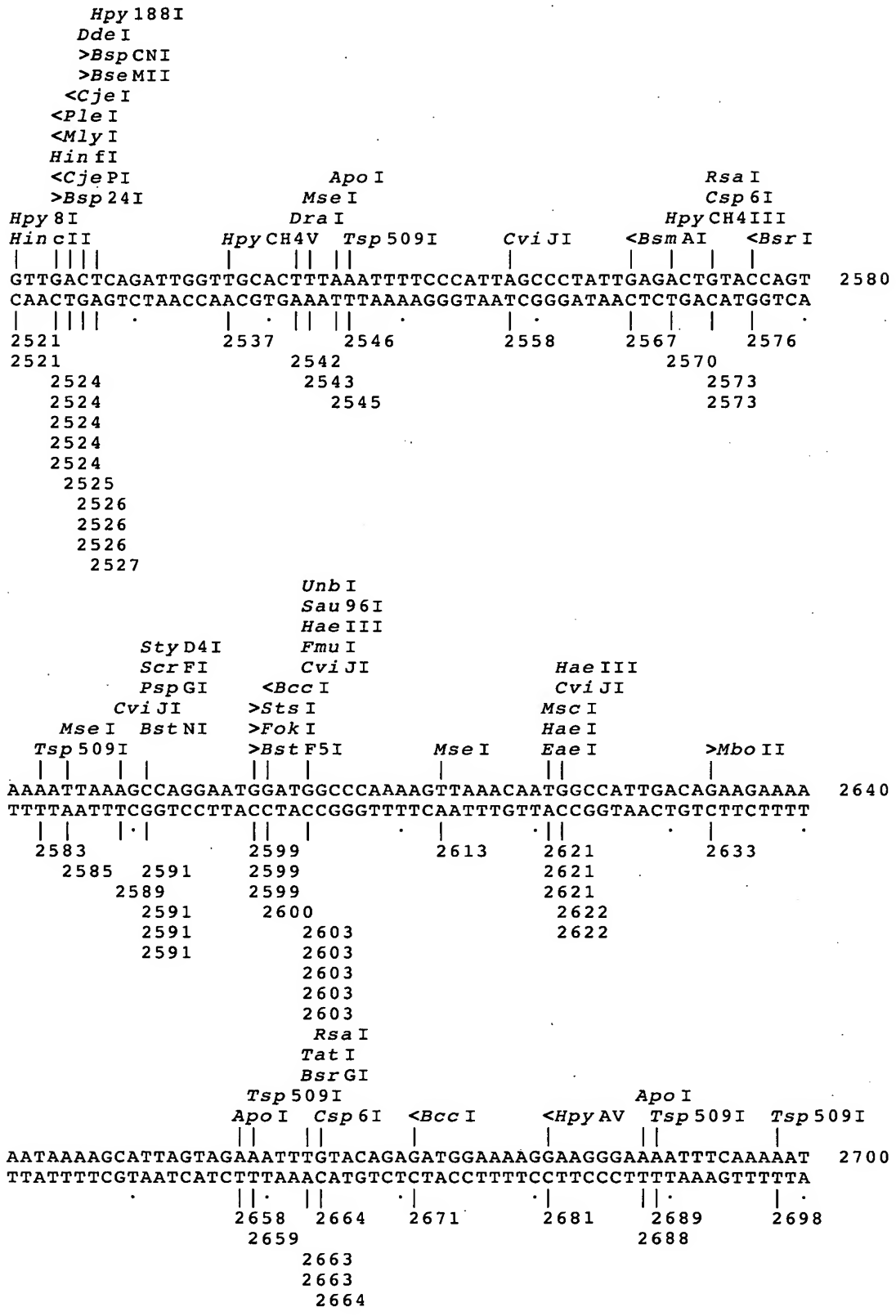

Vpa K11AI
Unb I
Sau 96I
Psp 03I
Mwo I
Cvi JI
<Bsm I
Bst XI
Hpy 188III
Fmu I
Ava II
>Fal I
Nla IV
<Bsm AI
Hpy 8I
Acc I
ATAGCCCTACCAGCATTCTGGACATAAGACAAGGACCAAAGGAACCCTTTAGAGACTATG 1680
TATCGGGATGGTCGTAAGACCTGTATTCTGTTCCTGGTTTCCTTGGGAAATCTCTGATAC
1623 1630 1637 1653 1661 1672 1680
1624 1633 1653 1659 1680
1653
1653
1653
1653
Hpa II
Bsr FI
Bsa WI
Age I
Dde I
Cvi JI
Mwo I
Cac 8I
Alu I
Hin dIII
<Mnl I
Cvi JI
Tsp 509I
>Sts I
>Fok I
>Bst F5I
TAGACCGGTTCTATAAAACTCTAAGAGCCGAGCAAGCTTCACAGGAGGTAAAAAATTGGA 1740
ATCTGGCCAAGATATTTTGAGATTCTCGGCTCGTTCGAAGTGTCCTCCATTTTTTAACCT
1684 1701 1707 1714 1725 1734
1684 1706 1715 1738
1684 1712 1738
1685 1715 1738
Vpa K11AI
Unb I
Sau 96I
Psp 03I
Fmu I
Ava II
Mse I
Dra I
TGACAGAAACCTTGTTGGTCCAAAATGCGAACCCAGATTGTAAGACTATTTTAAAAGCAT 1800
ACTGTCTTTGGAACAACCAGGTTTTACGCTTGGGTCTAACATTCTGATAAAATTTTCGTA
1757 1790
1757 1791
1757
1757
1757
1757

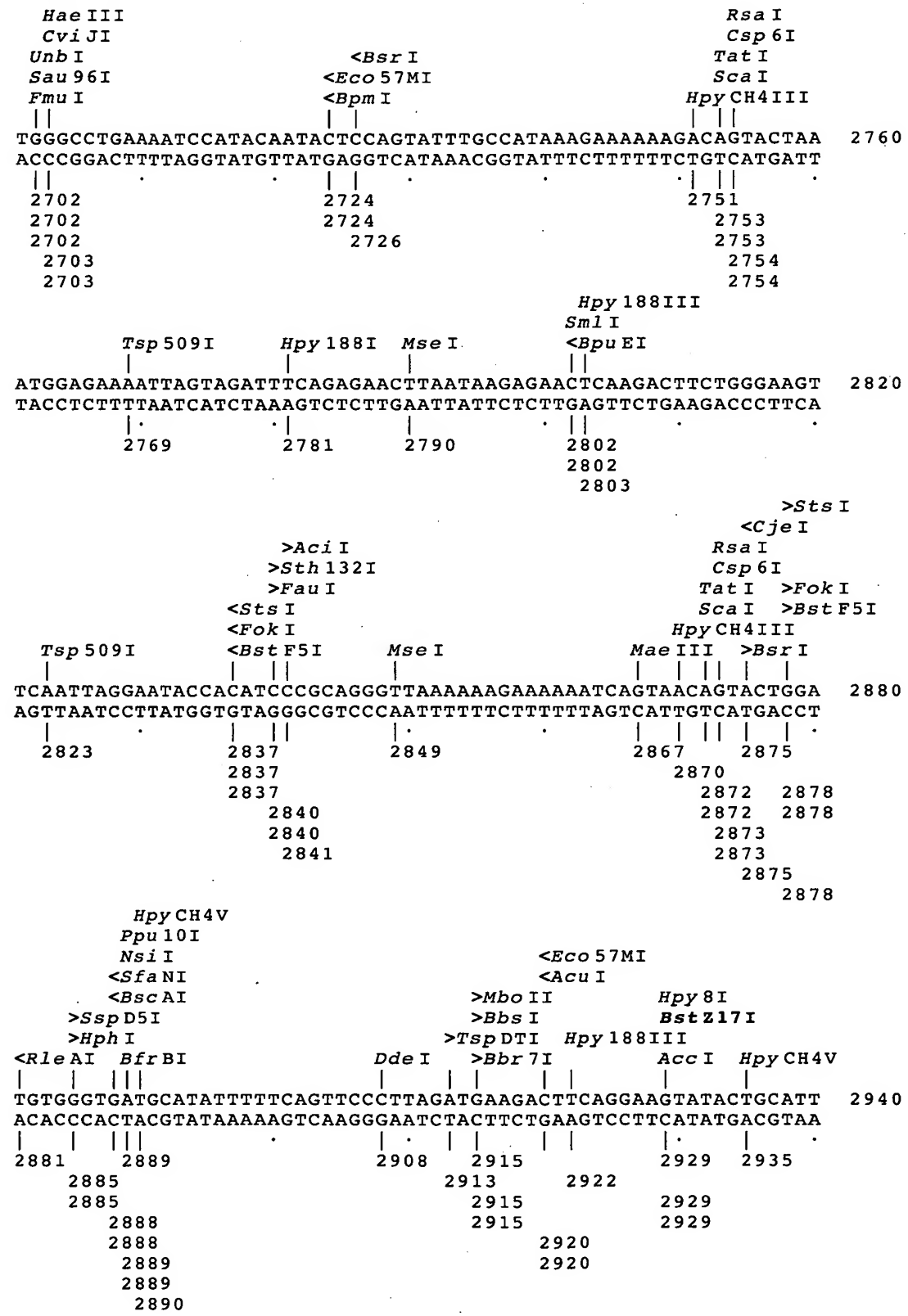

HpaII
StyD4I
>Sth132I
ScrFI
NciI
EcoHI
<SimI
VpaK11AI
UnbI
Sau96I
Psp03I
NlaIV
FmuI HaeIII
AvaII
PssI >GdiII
PpuMI CviJI
Eco0109I
<MnlI EaeI
TauI
MspA1I
VpaK11AI
UnbI FnuDII
Sau96I
Psp03I
FmuI BthCI
AvaII
NlaIV <AciI >MboII
>BsmFI CviJI BfaI
NlaIII
FatI
CviAII
NspI
>HinDIII
>HaeIV
AhdI
TGGGACCAGCGGCTACACTAGAAGAAATGATGACAGCATGTCAGGGAGTAGGAGGACCCG 1860
ACCCTGGTCGCCGATGTGATCTTCTTTACTACTGTCGTACAGTCCCTCATCCTCCTGGGC
1802 1811 1818 1832 1852 1859
1802 1809 1821 1832 1853
1803 1832 1853 1860
1803 1809 1836 1853 1859
1803 1837 1854
1803 1837 1854 1860
1803 1809 1837 1854
1803 1854
1807 1854
1809 1854
1854
1855
1857
1857
1857
1857
1857
1858
>Eco57MI
>AcuI CviJI ApoI CviJI
CviJI >BsrDI MaeIII AluI MslI
TspDTI
GCCATAAGGCAAGAGTTTTGGCTGAAGCAATGAGCCAAGTAACAAATTCAGCTACCATAA 1920
CGGTATTCCGTTCTCAAAACCGACTTCGTTACTCGGTTCATTGTTTAAGTCGATGGTATT
1880 1887 1899 1910 1916
1882 1893 1904 1910
1882 1905
HpyCH4V
<SfaNI TspDTI TspDTI
<BscAI <MnlI NlaIV MseI MfeI
TGATGCAGAGAGGCAATTTTAGGAACCAAAGAAAGATTGTTAAGTGTTTCAATTGTGGCA 1980
ACTACGTCTCTCCGTTAAAATCCTTGGTTTCTTTCTAACAATTCACAAAGTTAACACCGT
1922 1930 1942 1960 1970
1922 1935 1971
1924

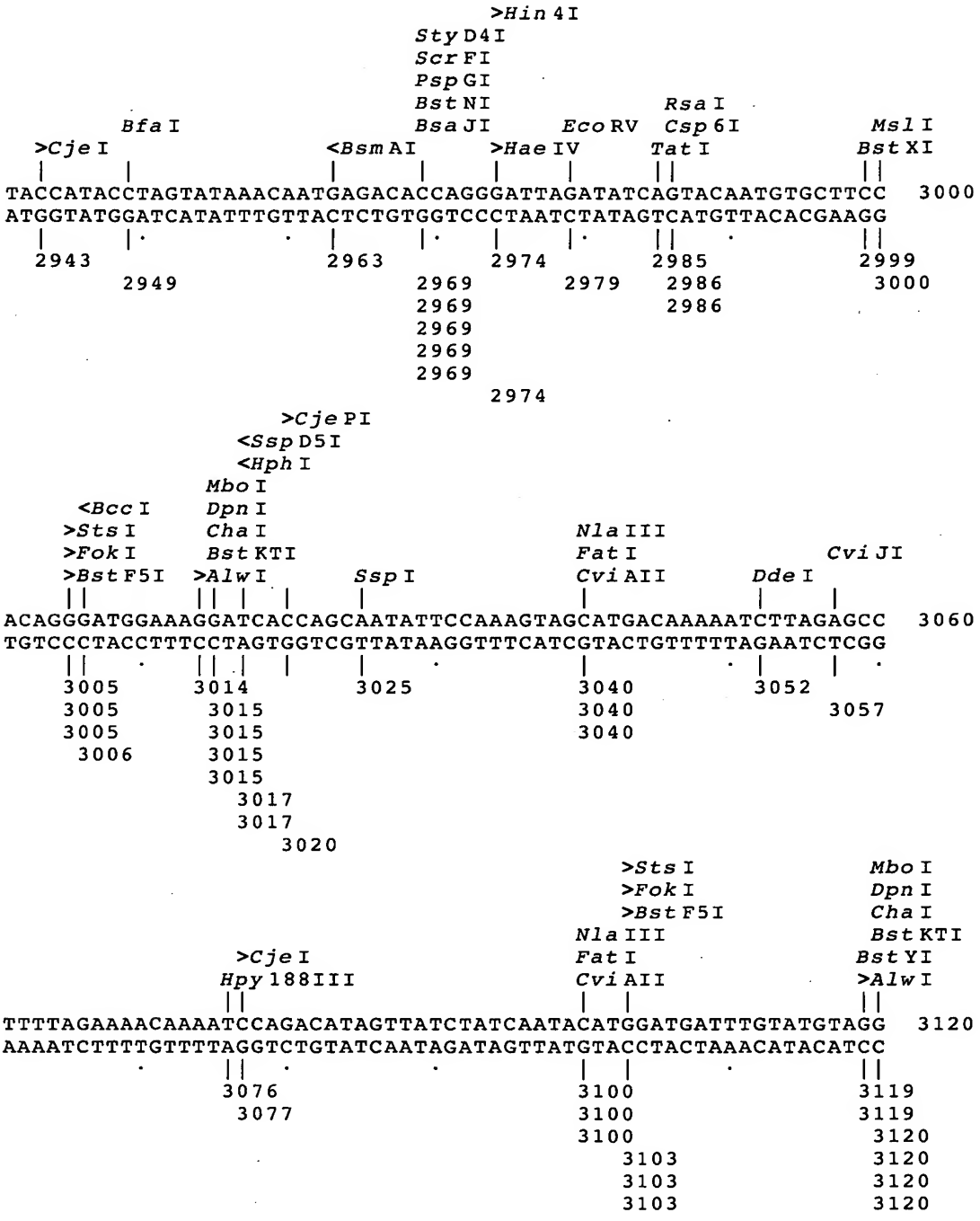

Unb I
Sau 96I
Nla IV
Hae III
Fmu I
Cvi JI
Unb I
Sau 96I
Pss I
PspOMI
Nla IV
Fmu I
Eco O109I
Bsp 1286I
Bme 1580I
Ban II
Apa I Bfa I
Pss I Sty I
Eco O109I
Bsp 1286I
Bme 1580I
Hpy CH4V Bsa JI
<Mme I
<Hpy AV Cvi JI Tsp 509I Avr II Cvi JI
AAGAAGGGCACACAGCCAGAAATTGCAGGGCCCCTAGGAAAAAGGGCTGTTGGAAATGTG 2040
TTCTTCCCGTGTGTCGGTCTTTAACGTCCCGGGGATCCTTTTTCCCGACAACCTTTACAC
1983 1994 2001 2013 2025
1986 2004 2013 2029
1986 2007
2007 2013
2008 2014
2008
2008
2008
2008
2008
2008
2008
2008
2008
2008
2009
2009
2009
2009
2009
2009
Mbo I
Dpn I
Cha I
Bst KTI
Dde I
<Bsp CNI
<Bse MII
Rsa I
Csp 6I
Bst YI
Bgl II
Cvi JI
<Hpy AV >Tsp DTI Tat I <Bsm AI Tsp 509I >Mbo II
GAAAGGAAGGACACCAAATGAAAGATTGTACTGAGAGACAGGCTAATTTTTTAGGGAAGA 2100
CTTTCCTTCCTGTGGTTTACTTTCTAACATGACTCTCTGTCCGATTAAAAAATCCCTTCT
2046 2058 2067 2075 2085 2096
2068 2081 2098
2068 2098
2071 2099
2071 2099
2071 2099
2099

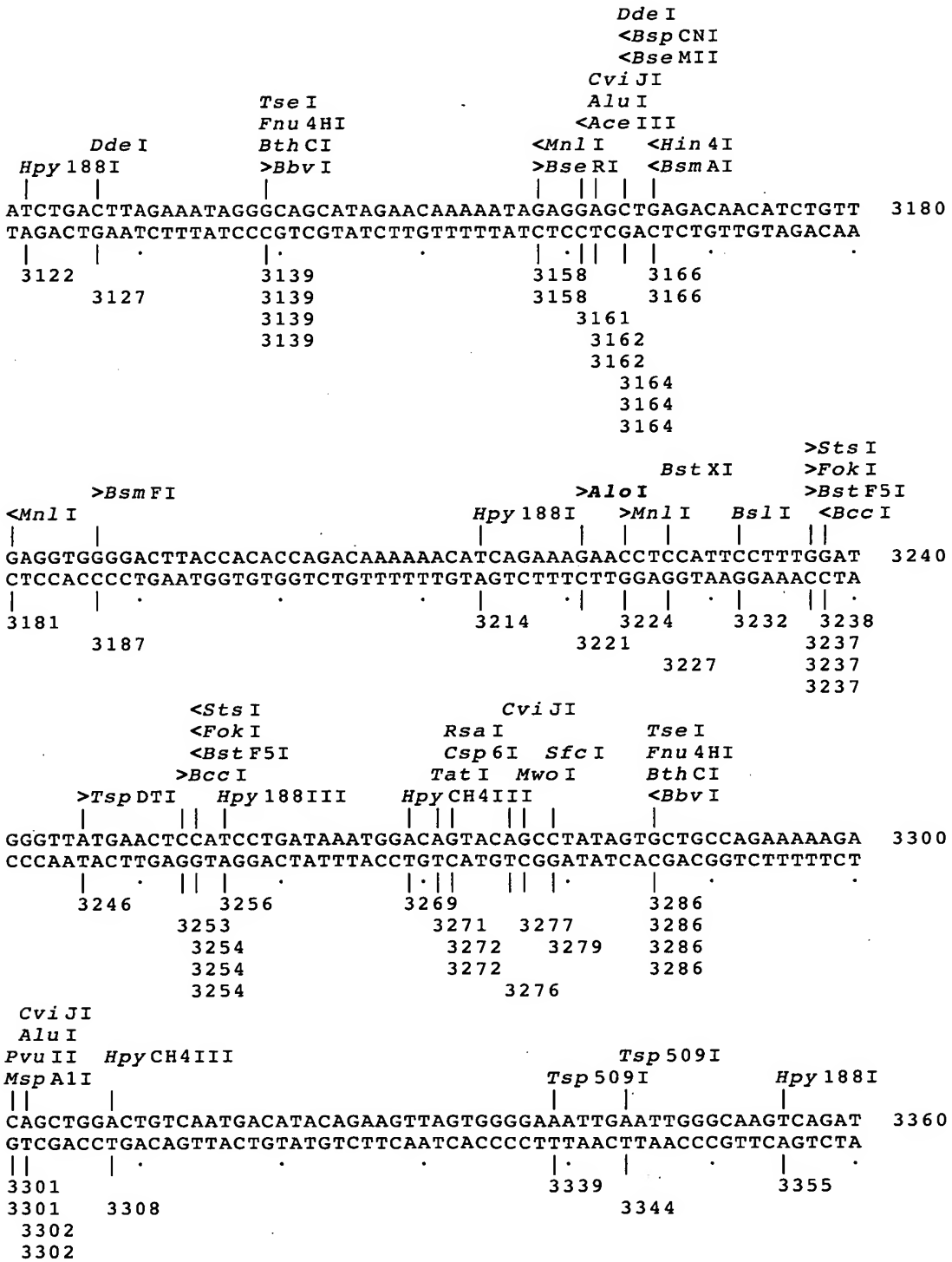

StyD4I
ScrFI
PspGI
BstNI
BsaJI
Hpy188I
>HpyAV
HaeIII
CviJI
HaeIII
CviJI
HaeI
<HpyAV
HaeI
CviJI
TspRI
ApoI
<MboII
CviJI
CviJI
TCTGGCCTTCCTACAAGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAG 2160
AGACCGGAAGGATGTTCCCTTCCGGTCCCTTAAAAGAAGTCTCGTCTGGTCTCGGTTGTC
2103
2104
2104
2106
2119
2121
2122
2122
2124
2124
2124
2124
2124
2129
2130
2135
2136
2136
2138
2152
2159
Hpy188I
CviJI
AluI
>BsaXI
>BsaXI
DdeI
>BspCNI
>MboII
<EcoRI
<EcoRI
<AcuI
>Hin4I
<BsmAI
>MnlI
>BseMII
NlaIV
CCCCACCAGAAGAGAGCTTCAGGTCTGGGGTAGAGACAACAACTCCCCCTCAGAAGCAGG 2220
GGGGTGGTCTTCTCTCGAAGTCCAGACCCCATCTCTGTTGTTGAGGGGGAGTCTTCGTCC
2169
2169
2175
2175
2177
2177
2193
2195
2196
2196
2208
2209
2209
2209
2210
2219
Tsp45I
MaeIII
DdeI
>BspCNI
>BseMII
<SimI
>Hin4I
>HaeIV
MseI
>BciVI
>MnlI
CviJI
HpyCH4III
Bsu36I
AhdI
AGCCGATAGACAAGGAACTGTATCCTTTAACTTCCCTCAGGTCACTCTTTGGCAACGACC 2280
TCGGCTATCTGTTCCTTGACATAGGAAATTGAAGGGAGTCCAGTGAGAAACCGTTGCTGG
2221
2237
2240
2247
2255
2255
2256
2256
2256
2261
2261
2277
2277
2277
2277
Tsp45I
MaeIII
>MnlI
CviJI
AluI
CCTCGTCACAATAAAGATAGGGGGGCAACTAAAGGAAGCTCTATTAGATACAGGAGCAGA 2340
GGAGCAGTGTTATTTCTATCCCCCCGTTGATTTCCTTCGAGATAATCTATGTCCTCGTCT
2281
2285
2285
2317
2317

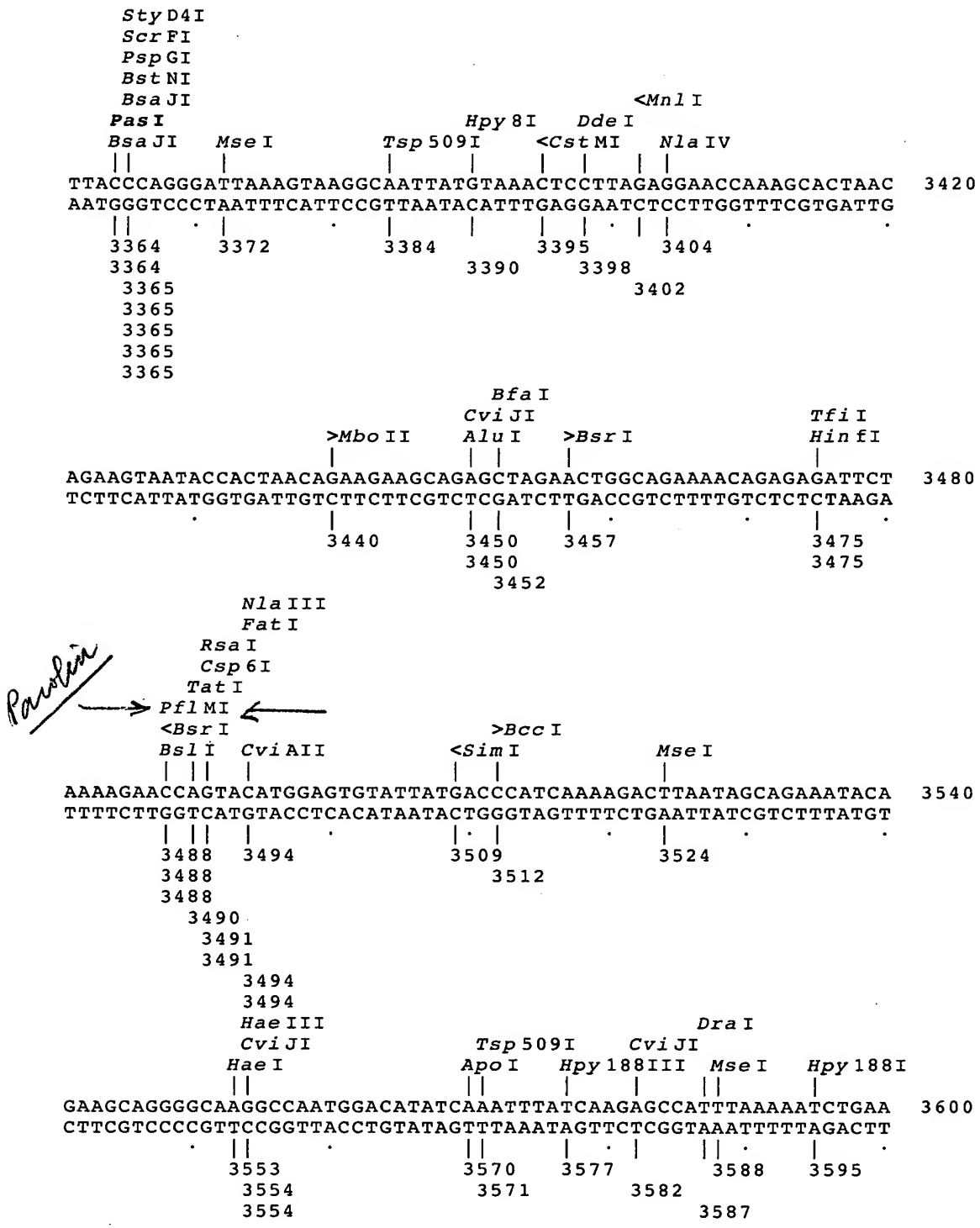

>Mbo II
Sty D4I
Scr FI
Psp GI
Bst XI
Bst NI
<Bcc I
Hpy CH4III
>Mbo II
TGATACAGTATTAGAAGAAATGAGTTTGCCAGGAAGATGGAAACCAAAAATGATAGGGGG 2400
ACTATGTCATAATCTTCTTTACTCAAACGGTCCTTCTACCTTTGGTTTTACTATCCCCC
2345
2354
2369
2376
2369
2369
2369
2369
2373
Hpy 188I
Mbo I
Dpn I
Cha I
Bst KTI
Bcl I
<Mnl I
<Hin 4I
Tsp 509I
Hpy CH4III
Hpy 8I
AATTGGAGGTTTTATCAAAGTAAGACAGTATGATCAGATACTCATAGAAATCTGTGGACA 2460
TTAACCTCCAAAATAGTTTCATTCTGTCATACTAGTCTATGAGTATCTTTAGACACCTGT
2401
2425
2454
2406
2406
2431
2432
2432
2432
2432
2434
Vpa K11AI
Unb I
Sau 96I
Psp 03I
Fmu I
Ava II
Sse 8647I
Sfc I
Hpy CH4III
Pss I
Cvi JI
Alu I
Rsa I
Csp 6I
Ppu MI
Eco O109I
Hpy 8I
Hin cII
Tsp 509I
>Mbo II
TAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAATCT 2520
ATTTCGATATCCATGTCATAATCATCCTGGATGTGGACAGTTGTATTAACCTTCTTTAGA
2464
2464
2466
2472
2472
2474
2485
2485
2485
2485
2486
2486
2486
2486
2486
2486
2498
2498
2506
2511

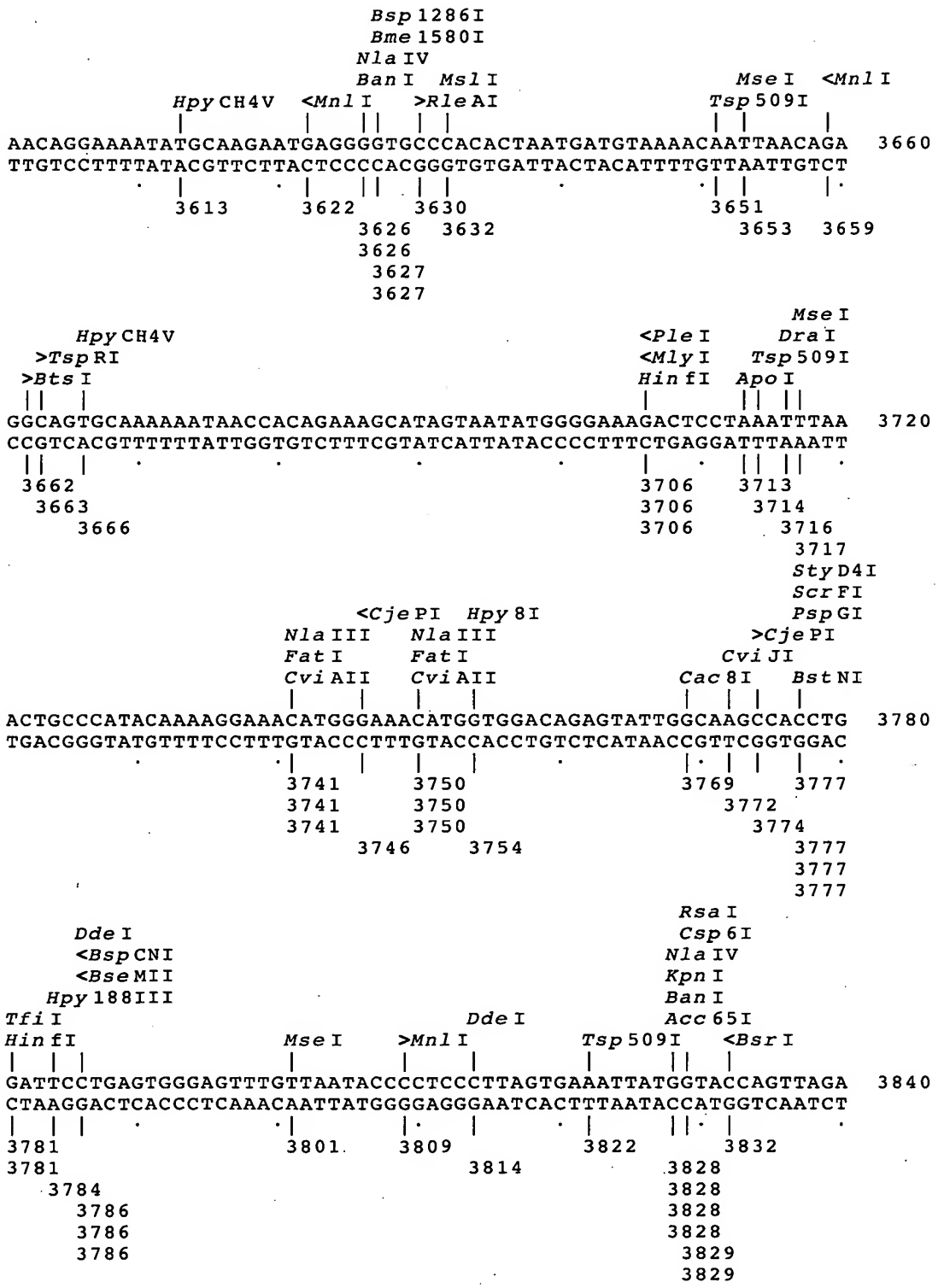

Hpy188I
DdeI
>BspCNI
>BseMII
<CjeI
<PleI
<MlyI
HinfI
<CjePI
>Bsp24I
ApoI
MseI
DraI
RsaI
Csp6I
HpyCH4III
Hpy8I
HincII
HpyCH4V
Tsp509I
CviJI
<BsmAI
<BsrI
GTTGACTCAGATTGGTTGCACTTTAAATTTTCCCATTAGCCCTATTGAGACTGTACCAGT 2580
CAACTGAGTCTAACCAACGTGAAATTTAAAAGGGTAATCGGGATAACTCTGACATGGTCA
2521 2537 2546 2558 2567 2576
2521 2542 2570
2524 2543 2573
2524 2545 2573
2524
2524
2524
2525
2526
2526
2526
2527
UnbI
Sau96I
HaeIII
StyD4I
FmuI
ScrFI
CviJI
HaeIII
PspGI
<BccI
CviJI
CviJI
>StsI
MscI
MseI
BstNI
>FokI
HaeI
Tsp509I
>BstF5I
MseI
EaeI
>MboII
AAAATTAAAGCCAGGAATGGATGGCCCAAAAGTTAAACAATGGCCATTGACAGAAGAAAA 2640
TTTTAATTTCGGTCCTTACCTACCGGGTTTTCAATTTGTTACCGGTAACTGTCTTCTTTT
2583 2599 2613 2621 2633
2585 2591 2599 2621
2589 2599 2621
2591 2600 2622
2591 2603 2622
2591 2603
2603
2603
2603
RsaI
TatI
BsrGI
Tsp509I
ApoI
ApoI
Csp6I
<BccI
<HpyAV
Tsp509I
Tsp509I
AATAAAAGCATTAGTAGAAATTTGTACAGAGATGGAAAAGGAAGGGAAAATTTCAAAAAT 2700
TTATTTTCGTAATCATCTTTAAACATGTCTCTACCTTTTCCTTCCCTTTTAAAGTTTTTA
2658 2664 2671 2681 2689 2698
2659 2688
2663
2663
2664

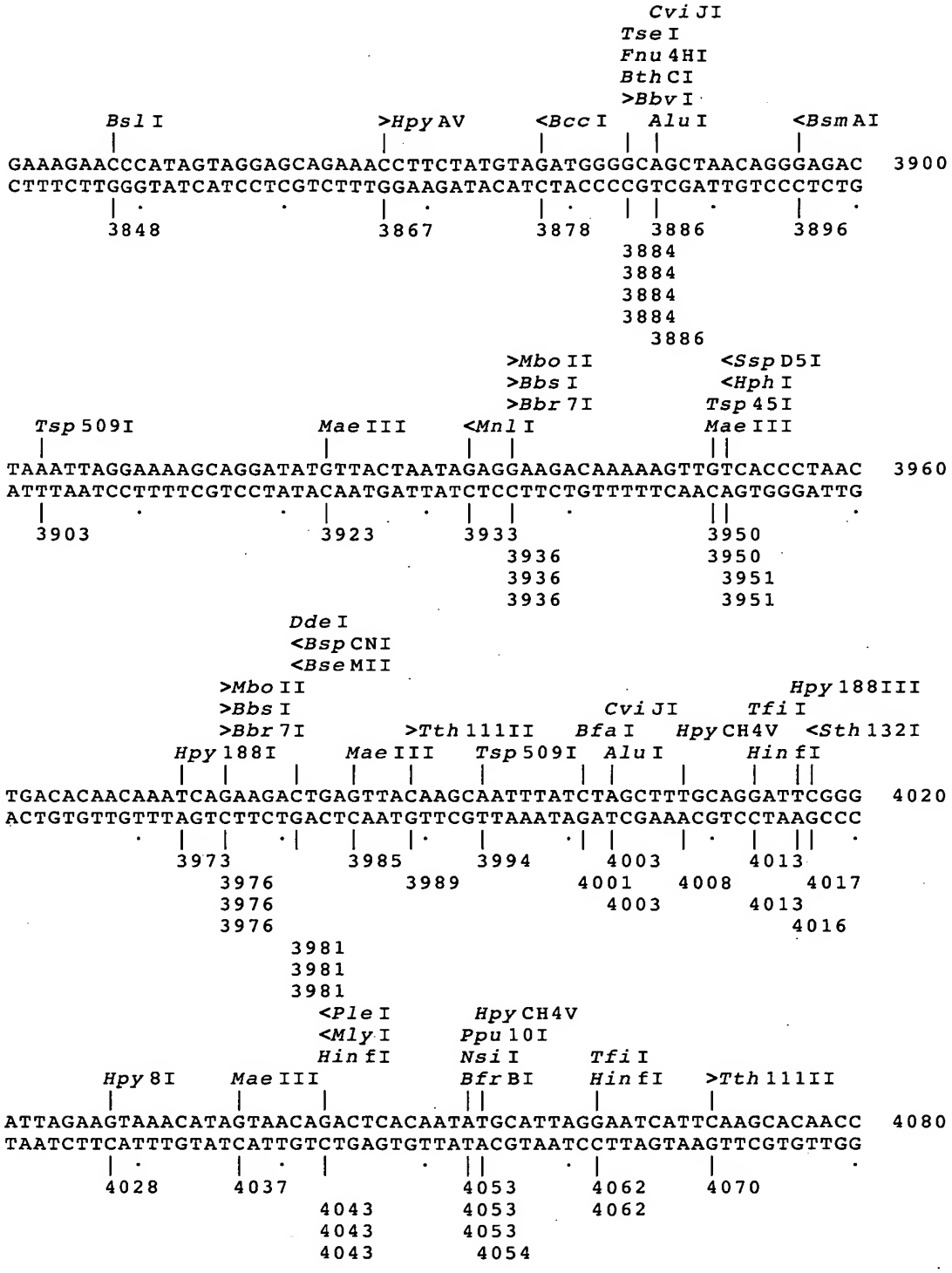

HaeIII
CviJI
UnbI
Sau96I
FmuI
<BsrI
<Eco57MI
<BpmI
RsaI
Csp6I
TatI
ScaI
HpyCH4III
TGGGCCTGAAAATCCATACAATACTCCAGTATTTGCCATAAAGAAAAAAGACAGTACTAA 2760
ACCCGGACTTTTAGGTATGTTATGAGGTCATAAACGGTATTTCTTTTTTCTGTCATGATT
2702
2702
2702
2703
2703
2724
2724
2726
2751
2753
2753
2754
2754
Hpy188III
SmlI
<BpuEI
Tsp509I
Hpy188I
MseI
ATGGAGAAAATTAGTAGATTTCAGAGAACTTAATAAGAGAACTCAAGACTTCTGGGAAGT 2820
TACCTCTTTTAATCATCTAAAGTCTCTTGAATTATTCTCTTGAGTTCTGAAGACCCTTCA
2769
2781
2790
2802
2802
2803
>StsI
<CjeI
>AciI
>Sth132I
>FauI
<StsI
<FokI
<BstF5I
RsaI
Csp6I
TatI
ScaI
>FokI
>BstF5I
HpyCH4III
Tsp509I
MseI
MaeIII
>BsrI
TCAATTAGGAATACCACATCCCGCAGGGTTAAAAAAGAAAAAATCAGTAACAGTACTGGA 2880
AGTTAATCCTTATGGTGTAGGGCGTCCCAATTTTTTCTTTTTTAGTCATTGTCATGACCT
2823
2837
2837
2837
2840
2840
2841
2849
2867
2870
2872
2872
2873
2873
2875
2875
2878
2878
HpyCH4V
Ppu10I
NsiI
<SfaNI
<BscAI
>SspD5I
>HphI
<Eco57MI
<AcuI
>MboII
>BbsI
>TspDTI
Hpy188III
Hpy8I
BstZ17I
<RleAI
BfrBI
DdeI
>Bbr7I
AccI
HpyCH4V
TGTGGGTGATGCATATTTTTCAGTTCCCTTAGATGAAGACTTCAGGAAGTATACTGCATT 2940
ACACCCACTACGTATAAAAAGTCAAGGGAATCTACTTCTGAAGTCCTTCATATGACGTAA
2881
2885
2885
2888
2888
2889
2889
2890
2889
2908
2913
2915
2915
2915
2920
2920
2922
2929
2929
2929
2935

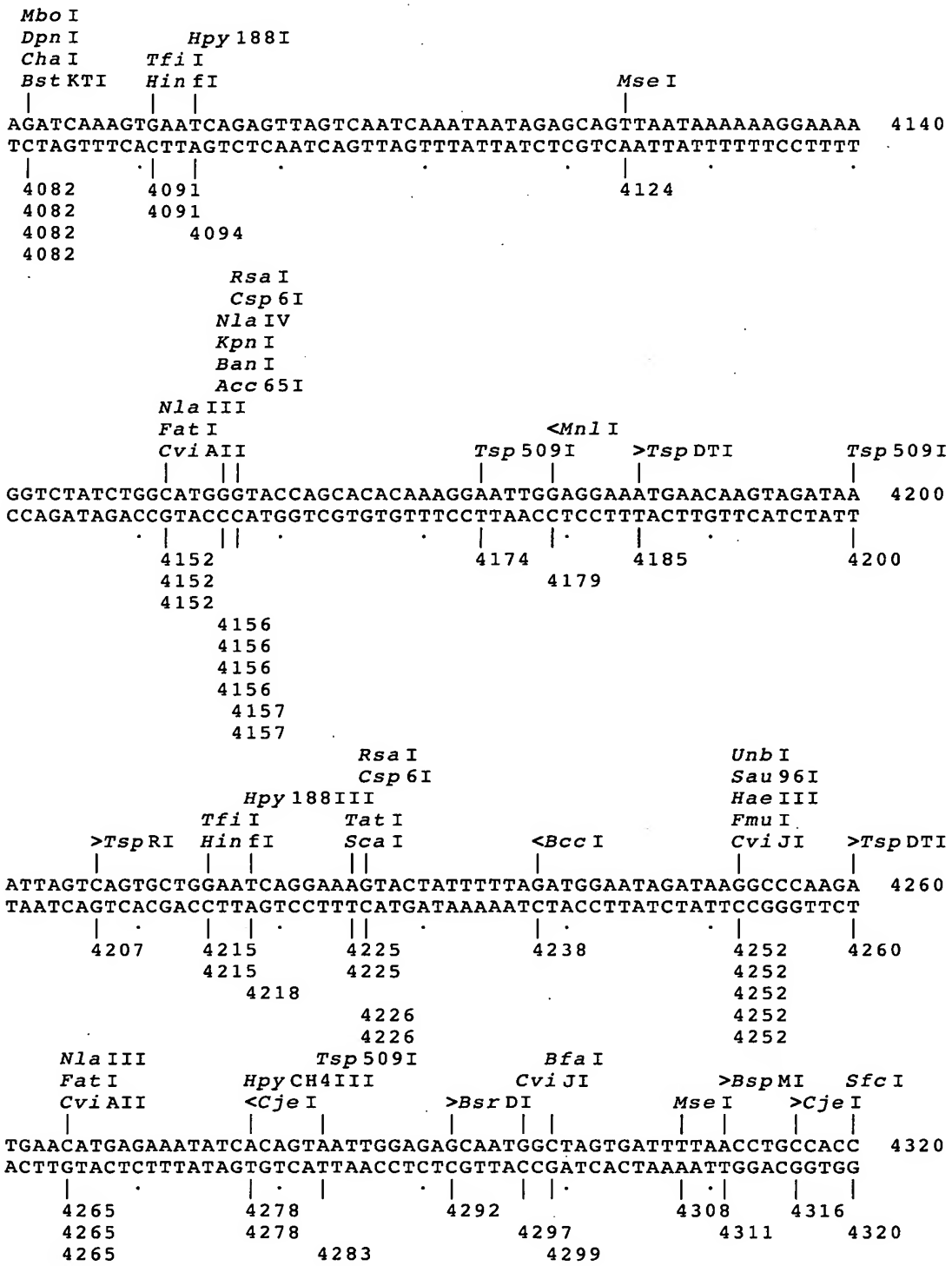

>Hin4I
StyD4I
ScrFI
PspGI
BstNI
RsaI
BfaI
BsaJI
EcoRV
Csp6I
MslI
>CjeI
<BsmAI
>HaeIV
TatI
BstXI
TACCATACCTAGTATAAACAATGAGACACCAGGGATTAGATATCAGTACAATGTGCTTCC 3000
ATGGTATGGATCATATTTGTTACTCTGTGGTCCCTAATCTATAGTCATGTTACACGAAGG
2943 2949 2963 2969 2969 2969 2969 2969 2974 2974 2979 2985 2986 2986 2999 3000
>CjePI
<SspD5I
<HphI
MboI
<BccI
DpnI
>StsI
ChaI
NlaIII
>FokI
BstKTI
FatI
CviJI
>BstF5I
>AlwI
SspI
CviAII
DdeI
ACAGGGATGGAAAGGATCACCAGCAATATTCCAAAGTAGCATGACAAAAATCTTAGAGCC 3060
TGTCCCTACCTTTCCTAGTGGTCGTTATAAGGTTTCATCGTACTGTTTTTAGAATCTCGG
3005 3005 3005 3006 3014 3015 3015 3015 3017 3017 3020 3025 3040 3040 3040 3052 3057
>StsI
>FokI
>BstF5I
MboI
DpnI
ChaI
BstKTI
NlaIII
FatI
BstYI
>CjeI
>AlwI
Hpy188III
CviAII
TTTTAGAAAACAAAATCCAGACATAGTTATCTATCAATACATGGATGATTTGTATGTAGG 3120
AAAATCTTTTGTTTTAGGTCTGTATCAATAGATAGTTATGTACCTACTAAACATACATCC
3076 3077 3100 3100 3100 3103 3103 3103 3119 3119 3120 3120 3120 3120

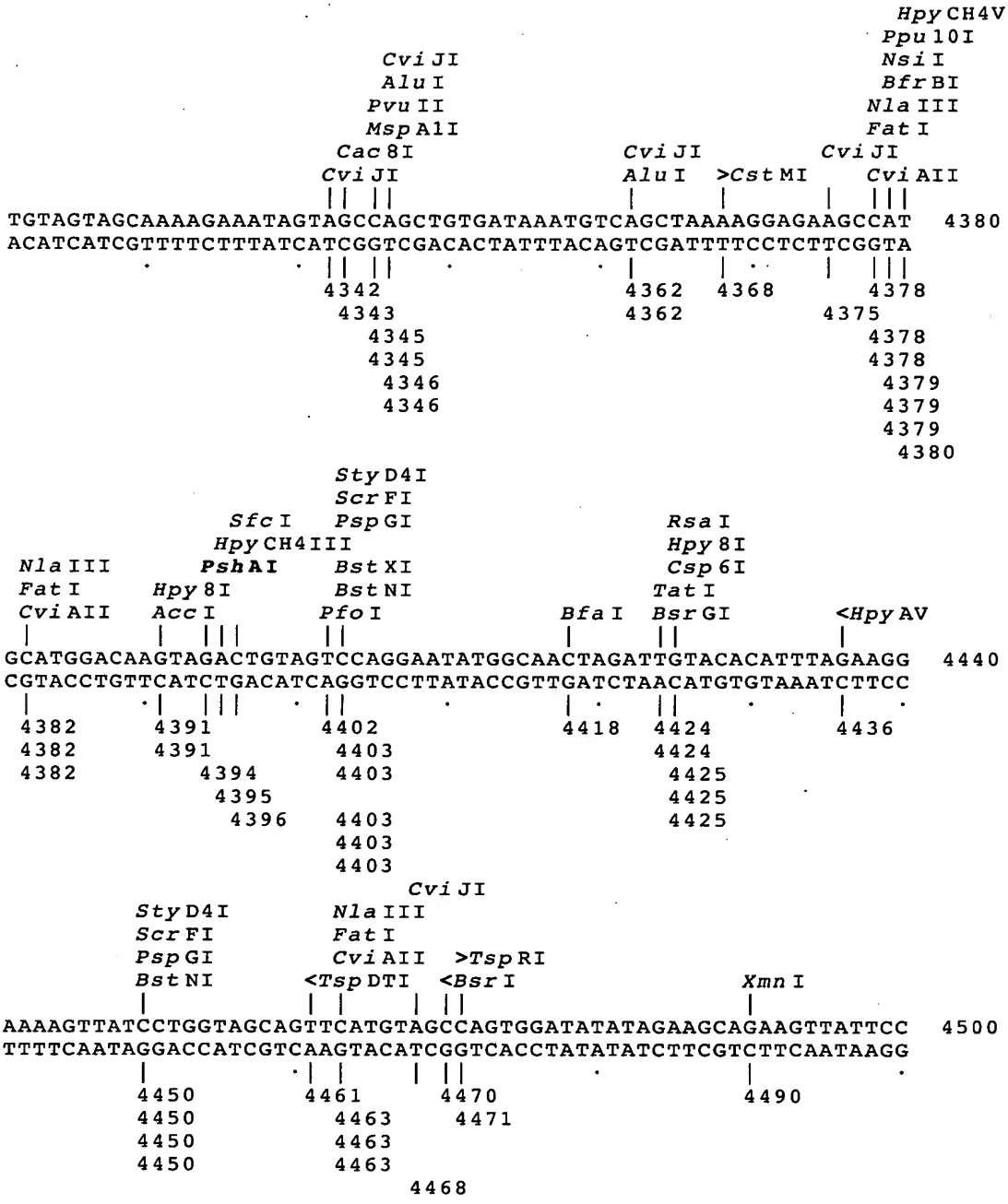

Dde I
<Bsp CNI
<Bse MII
Cvi JI
Tse I
Alu I
Fnu 4HI
<Ace III
Dde I
Bth CI
<Mnl I
<Hin 4I
Hpy 188I
>Bbv I
>Bse RI
<Bsm AI
ATCTGACTTAGAAATAGGGCAGCATAGAACAAAAATAGAGGAGCTGAGACAACATCTGTT 3180
TAGACTGAATCTTTATCCCGTCGTATCTTGTTTTTATCTCCTCGACTCTGTTGTAGACAA
3122
3127
3139
3139
3139
3139
3158
3158
3161
3162
3162
3164
3164
3164
>Sts I
Bst XI
>Fok I
>Bsm FI
>Alo I
>Bst F5I
<Mnl I
Hpy 188I
>Mnl I
Bsl I
<Bcc I
GAGGTGGGGACTTACCACACCAGACAAAAAACATCAGAAAGAACCTCCATTCCTTTGGAT 3240
CTCCACCCCTGAATGGTGTGGTCTGTTTTTTGTAGTCTTTCTTGGAGGTAAGGAAACCTA
3181
3187
3214
3221
3224
3227
3232
3237
3237
3237
<Sts I
<Fok I
<Bst F5I
>Bcc I
Cvi JI
Rsa I
Csp 6I
Tat I
Sfc I
Mwo I
Tse I
Fnu 4HI
Bth CI
<Bbv I
>Tsp DTI
Hpy 188III
Hpy CH4III
GGGTTATGAACTCCATCCTGATAAATGGACAGTACAGCCTATAGTGCTGCCAGAAAAAGA 3300
CCCAATACTTGAGGTAGGACTATTTACCTGTCATGTCGGATATCACGACGGTCTTTTTCT
3246
3253
3254
3254
3254
3256
3269
3271
3272
3272
3276
3277
3279
3286
3286
3286
3286
Cvi JI
Alu I
Pvu II
Hpy CH4III
Tsp 509I
Msp A1I
Tsp 509I
Hpy 188I
CAGCTGGACTGTCAATGACATACAGAAGTTAGTGGGGAAATTGAATTGGGCAAGTCAGAT 3360
GTCGACCTGACAGTTACTGTATGTCTTCAATCACCCCTTTAACTTAACCCGTTCAGTCTA
3301
3301
3302
3302
3308
3339
3344
3355

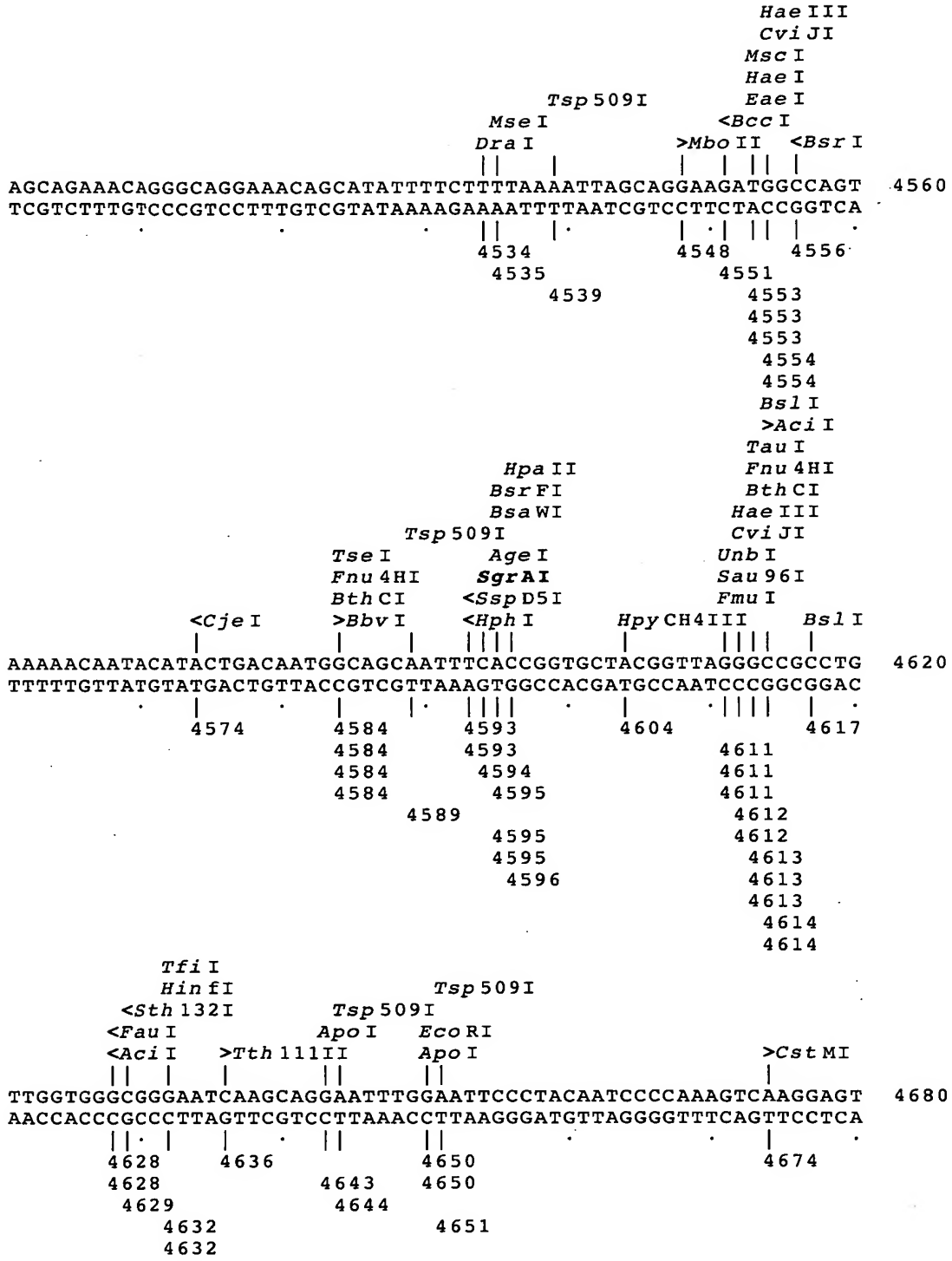

Sty D4I
Scr FI
Psp GI
Bst NI
Bsa JI
<Mnl I
Pas I
Hpy 8I
Dde I
Bsa JI
Mse I
Tsp 509I
<Cst MI
Nla IV
TTACCCAGGGATTAAAGTAAGGCAATTATGTAAACTCCTTAGAGGAACCAAAGCACTAAC 3420
AATGGGTCCCTAATTTCATTCCGTTAATACATTTGAGGAATCTCCTTGGTTTCGTGATTG
3364 3372 3384 3395 3404
3364 3390 3398
3365 3402
3365
3365
3365
3365
Bfa I
Cvi JI
>Mbo II
Alu I
>Bsr I
Tfi I
Hin fI
AGAAGTAATACCACTAACAGAAGAAGCAGAGCTAGAACTGGCAGAAAACAGAGAGATTCT 3480
TCTTCATTATGGTGATTGTCTTCTTCGTCTCGATCTTGACCGTCTTTTGTCTCTCTAAGA
3440 3450 3457 3475
3450 3475
3452
Nla III
Fat I
Rsa I
Csp 6I
Tat I
Pfl MI
<Bsr I
>Bcc I
Bsl I
Cvi AII
<Sim I
Mse I
AAAAGAACCAGTACATGGAGTGTATTATGACCCATCAAAAGACTTAATAGCAGAAATACA 3540
TTTTCTTGGTCATGTACCTCACATAATACTGGGTAGTTTTCTGAATTATCGTCTTTATGT
3488 3494 3509 3524
3488 3512
3488
3490
3491
3491
3494
3494
Hae III
Dra I
Cvi JI
Tsp 509I
Cvi JI
Hae I
Apo I
Hpy 188III
Mse I
Hpy 188I
GAAGCAGGGGCAAGGCCAATGGACATATCAAATTTATCAAGAGCCATTTAAAAATCTGAA 3600
CTTCGTCCCCGTTCCGGTTACCTGTATAGTTTAAATAGTTCTCGGTAAATTTTTAGACTT
3553 3570 3577 3588 3595
3554 3571 3582
3554 3587

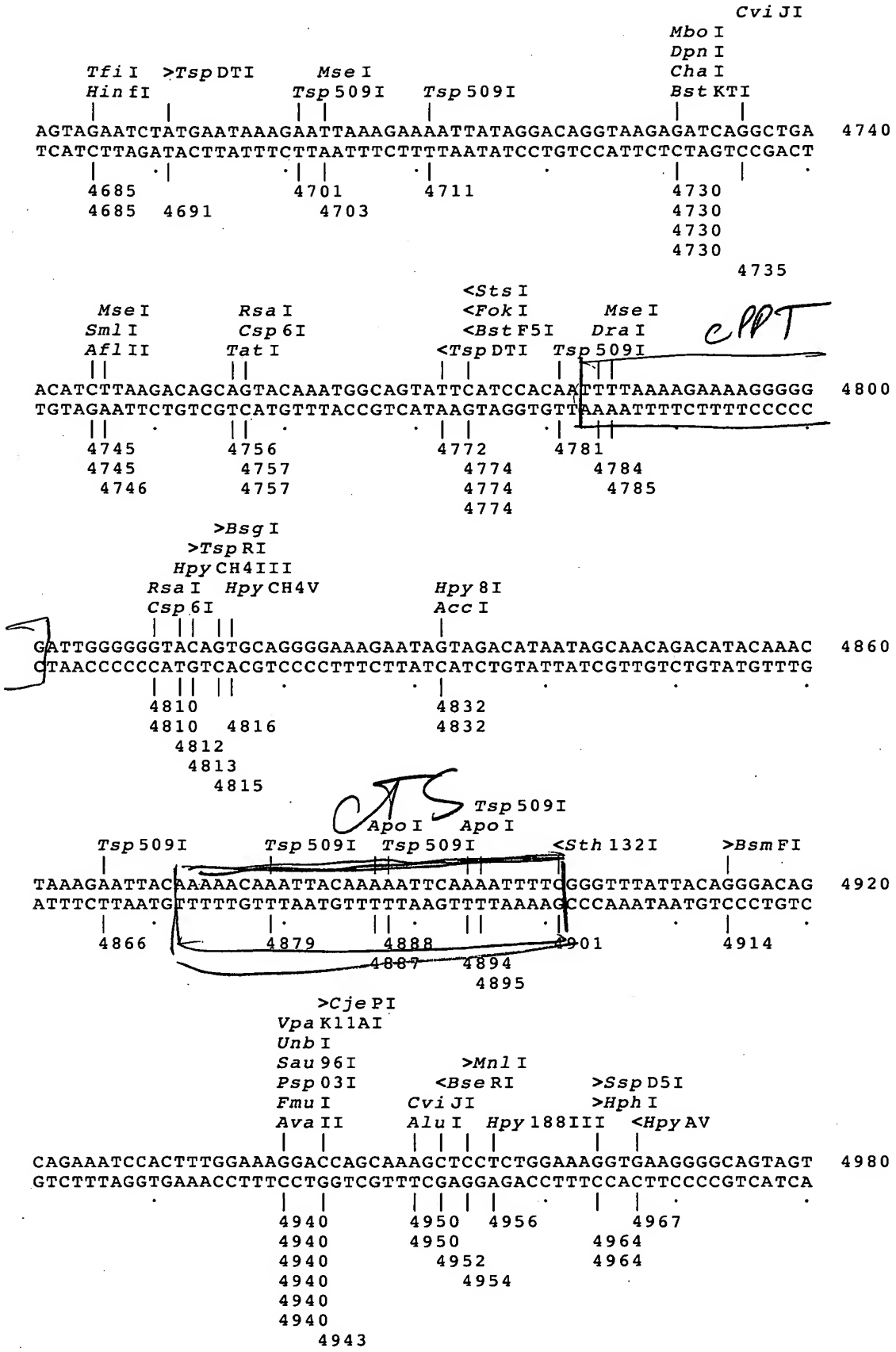

Bsp 1286I
Bme 1580I
Nla IV
Ban I Msl I
Mse I <Mnl I
Hpy CH4V <Mnl I >Rle AI
Tsp 509I
AACAGGAAAATATGCAAGAATGAGGGGTGCCCACACTAATGATGTAAAACAATTAACAGA 3660
TTGTCCTTTTATACGTTCTTACTCCCCACGGGTGTGATTACTACATTTTGTTAATTGTCT
3613 3622 3630 3651
3626 3632 3653 3659
3626
3627
3627
Mse I
Hpy CH4V <Ple I Dra I
>Tsp RI <Mly I Tsp 509I
>Bts I Hin fI Apo I
GGCAGTGCAAAAAATAACCACAGAAAGCATAGTAATATGGGGAAAGACTCCTAAATTTAA 3720
CCGTCACGTTTTTTATTGGTGTCTTTCGTATCATTATACCCCTTTCTGAGGATTTAAATT
3662 3706 3713
3663 3706 3714
3666 3706 3716
3717
Sty D4I
Scr FI
Psp GI
<Cje PI Hpy 8I
Nla III Nla III >Cje PI
Fat I Fat I Cvi JI
Cvi AII Cvi AII Cac 8I Bst NI
ACTGCCCATACAAAAGGAAACATGGGAAACATGGTGGACAGAGTATTGGCAAGCCACCTG 3780
TGACGGGTATGTTTTCCTTTGTACCCTTTGTACCACCTGTCTCATAACCGTTCGGTGGAC
3741 3750 3769 3777
3741 3750 3772
3741 3750 3774
3746 3754 3777
3777
3777
Rsa I
Dde I Csp 6I
<Bsp CNI Nla IV
<Bse MII Kpn I
Hpy 188III Ban I
Tfi I Dde I Acc 65I
Hin fI Mse I >Mnl I Tsp 509I <Bsr I
GATTCCTGAGTGGGAGTTTGTTAATACCCCTCCCTTAGTGAAATTATGGTACCAGTTAGA 3840
CTAAGGACTCACCCTCAAACAATTATGGGGAGGGAATCACTTTAATACCATGGTCAATCT
3781 3801 3809 3822 3832
3781 3814 3828
3784 3828
3786 3828
3786 3828
3786 3829
3829

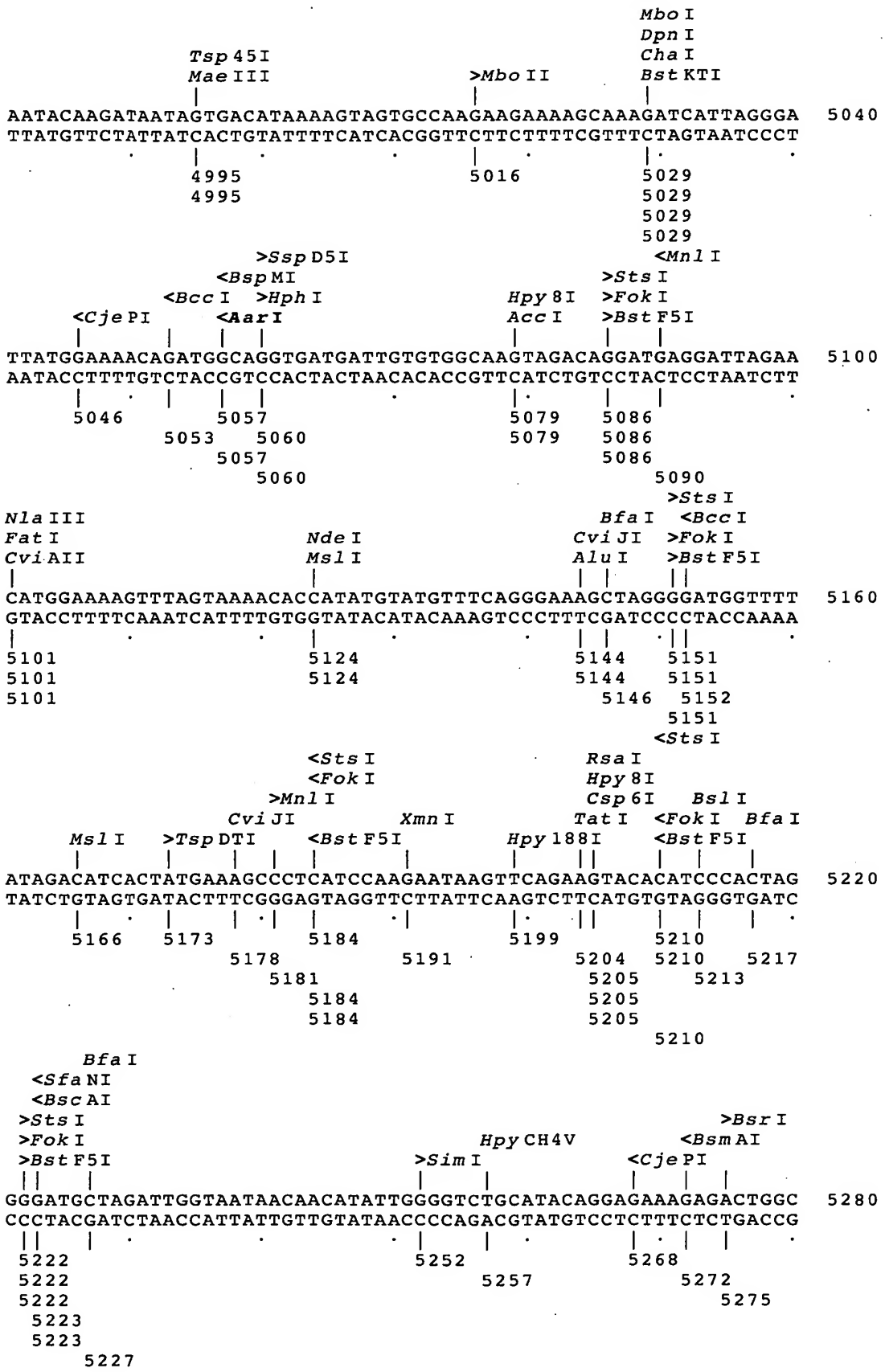

CviJI
TseI
Fnu4HI
BthCI
>BbvI
BslI
>HpyAV
<BccI
AluI
<BsmAI
GAAAGAACCCATAGTAGGAGCAGAAACCTTCTATGTAGATGGGGCAGCTAACAGGGAGAC 3900
CTTTCTTGGGTATCATCCTCGTCTTTGGAAGATACATCTACCCCGTCGATTGTCCCTCTG
3848 3867 3878 3886 3896
3884
3884
3884
3884
3886
>MboII
>BbsI
>Bbr7I
<SspD5I
<HphI
Tsp45I
Tsp509I
MaeIII
<MnlI
MaeIII
TAAATTAGGAAAAGCAGGATATGTTACTAATAGAGGAAGACAAAAAGTTGTCACCCTAAC 3960
ATTTAATCCTTTTCGTCCTATACAATGATTATCTCCTTCTGTTTTTCAACAGTGGGATTG
3903 3923 3933 3950
3936 3950
3936 3951
3936 3951
DdeI
<BspCNI
<BseMII
>MboII
>BbsI
>Bbr7I
>Tth111II
CviJI
BfaI
HpyCH4V
<Sth132I
Hpy188III
Hpy188I
MaeIII
Tsp509I
AluI
TfiI
HinfI
TGACACAACAAATCAGAAGACTGAGTTACAAGCAATTTATCTAGCTTTGCAGGATTCGGG 4020
ACTGTGTTGTTTAGTCTTCTGACTCAATGTTCGTTAAATAGATCGAAACGTCCTAAGCCC
3973 3985 3994 4003 4013
3976 3989 4001 4008 4017
3976 4003 4013
3976 4016
3981
3981
3981
<PleI
<MlyI
HinfI
HpyCH4V
Ppu10I
NsiI
TfiI
Hpy8I
MaeIII
BfrBI
HinfI
>Tth111II
ATTAGAAGTAAACATAGTAACAGACTCACAATATGCATTAGGAATCATTCAAGCACAACC 4080
TAATCTTCATTTGTATCATTGTCTGAGTGTTATACGTAATCCTTAGTAAGTTCGTGTTGG
4028 4037 4053 4062 4070
4043 4053 4062
4043 4053
4043 4054

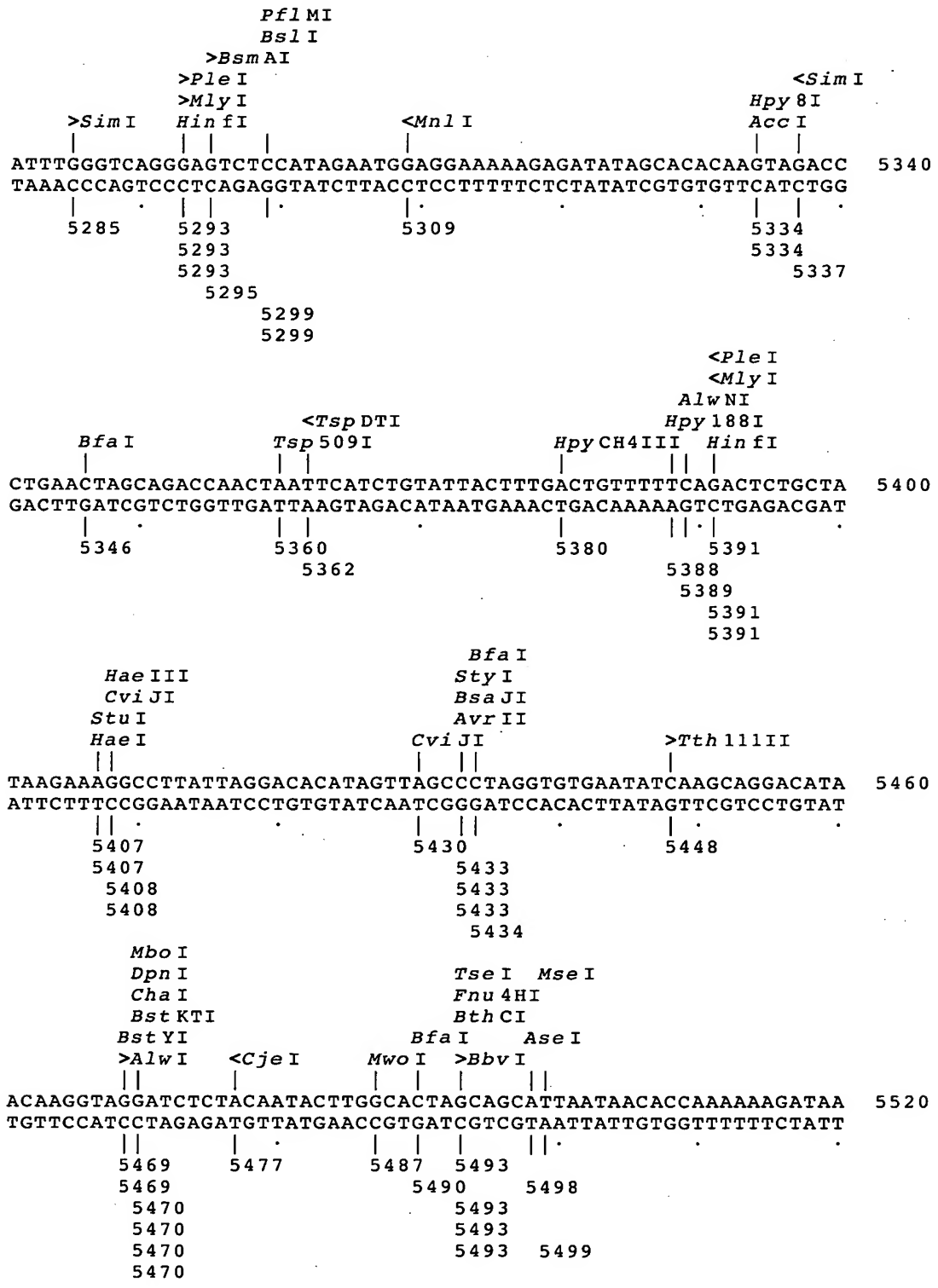

Mbo I
Dpn I
Hpy 188I
Cha I
Tfi I
Bst KTI
Hin fI
Mse I
AGATCAAAGTGAATCAGAGTTAGTCAATCAAATAATAGAGCAGTTAATAAAAAAGGAAAA 4140
TCTAGTTTCACTTAGTCTCAATCAGTTAGTTTATTATCTCGTCAATTATTTTTTCCTTTT
4082
4091
4124
4082
4091
4082
4094
4082
Rsa I
Csp 6I
Nla IV
Kpn I
Ban I
Acc 65I
Nla III
Fat I
Cvi AII
<Mnl I
Tsp 509I
>Tsp DTI
Tsp 509I
GGTCTATCTGGCATGGGTACCAGCACACAAAGGAATTGGAGGAAATGAACAAGTAGATAA 4200
CCAGATAGACCGTACCCATGGTCGTGTGTTTCCTTAACCTCCTTTACTTGTTCATCTATT
4152
4174
4185
4200
4152
4179
4152
4156
4156
4156
4156
4157
4157
Rsa I
Csp 6I
Unb I
Sau 96I
Hae III
Fmu I
Hpy 188III
Tfi I
Tat I
>Tsp RI
Hin fI
Sca I
<Bcc I
Cvi JI
>Tsp DTI
ATTAGTCAGTGCTGGAATCAGGAAAGTACTATTTTTAGATGGAATAGATAAGGCCCAAGA 4260
TAATCAGTCACGACCTTAGTCCTTTCATGATAAAAATCTACCTTATCTATTCCGGGTTCT
4207
4215
4225
4238
4252
4260
4215
4225
4252
4218
4252
4226
4252
4226
4252
Nla III
Fat I
Cvi AII
Tsp 509I
Hpy CH4III
<Cje I
Bfa I
Cvi JI
>Bsr DI
>Bsp MI
Mse I
Sfc I
>Cje I
TGAACATGAGAAATATCACAGTAATTGGAGAGCAATGGCTAGTGATTTTAACCTGCCACC 4320
ACTTGTACTCTTTATAGTGTCATTAACCTCTCGTTACCGATCACTAAAATTGGACGGTGG
4265
4278
4292
4308
4316
4265
4278
4297
4311
4320
4265
4283
4299

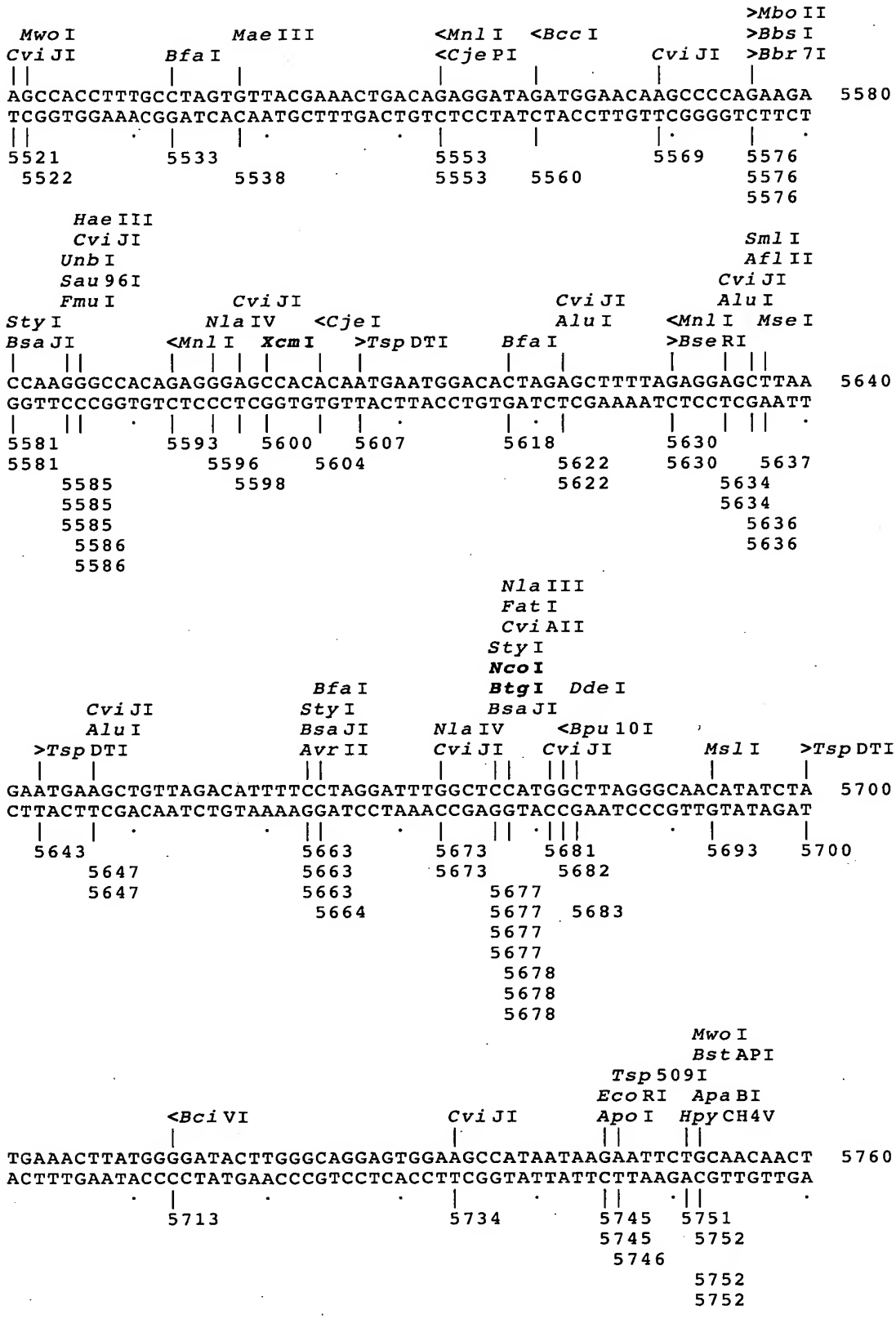

HpyCH4V
Ppu10I
CviJI
NsiI
AluI
BfrBI
PvuII
NlaIII
MspAlI
FatI
Cac8I
CviJI
CviJI
CviJI
AluI
>CstMI
CviAII
TGTAGTAGCAAAAGAAATAGTAGCCAGCTGTGATAAATGTCAGCTAAAAGGAGAAGCCAT 4380
ACATCATCGTTTTCTTTATCATCGGTCGACACTATTTACAGTCGATTTTCCTCTTCGGTA
4342
4343
4345
4345
4346
4346
4362
4362
4368
4375
4378
4378
4379
4379
4379
4380
StyD4I
ScrFI
SfcI
PspGI
RsaI
HpyCH4III
Hpy8I
NlaIII
PshAI
BstXI
CspI6I
FatI
Hpy8I
BstNI
TatI
CviAII
AccI
PfoI
BfaI
BsrGI
<HpyAV
GCATGGACAAGTAGACTGTAGTCCAGGAATATGGCAACTAGATTGTACACATTTAGAAGG 4440
CGTACCTGTTCATCTGACATCAGGTCCTTATACCGTTGATCTAACATGTGTAAATCTTCC
4382
4382
4382
4391
4391
4394
4395
4396
4402
4403
4403
4403
4403
4418
4424
4424
4425
4425
4425
4436
CviJI
StyD4I
NlaIII
ScrFI
FatI
PspGI
CviAII
>TspRI
BstNI
<TspDTI
<BsrI
XmnI
AAAAGTTATCCTGGTAGCAGTTCATGTAGCCAGTGGATATATAGAAGCAGAAGTTATTCC 4500
TTTTCAATAGGACCATCGTCAAGTACATCGGTCACCTATATATCTTCGTCTTCAATAAGG
4450
4450
4450
4450
4461
4463
4463
4463
4468
4470
4471
4490

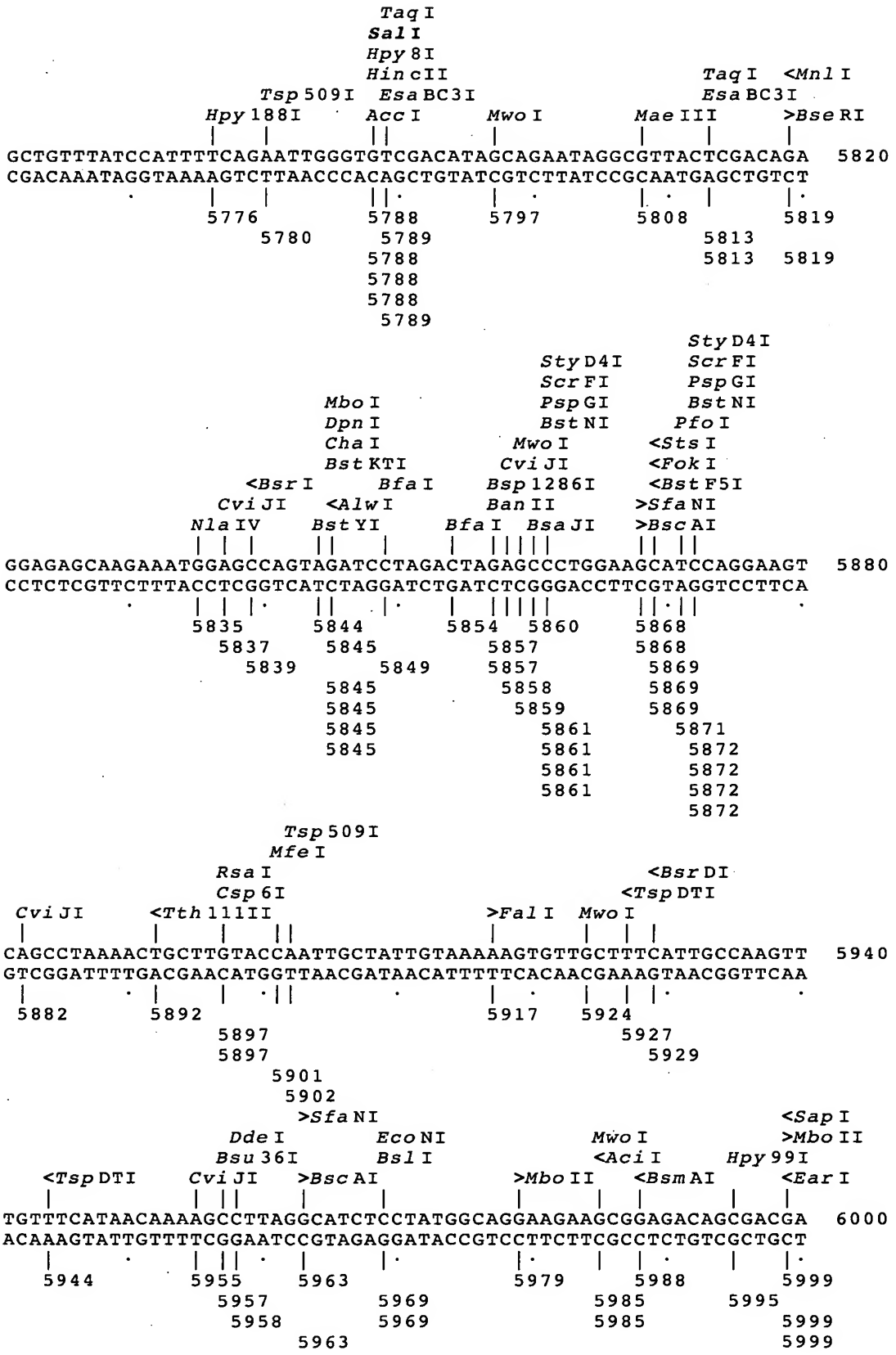

HaeIII
CviJI
MscI
HaeI
EaeI
Tsp509I
MseI
<BccI
DraI
>MboII
<BsrI
AGCAGAAACAGGGCAGGAAACAGCATATTTTCTTTTAAAATTAGCAGGAAGATGGCCAGT 4560
TCGTCTTTGTCCCGTCCTTTGTCGTATAAAAGAAAATTTTAATCGTCCTTCTACCGGTCA
4534 4548 4556
4535 4551
4539 4553
4553
4553
4554
4554
BslI
>AciI
TauI
HpaII
FnuHI
BsrFI
BthCI
BsaWI
HaeIII
Tsp509I
CviJI
TseI
AgeI
UnbI
FnuHI
SgrAI
Sau96I
BthCI
<SspD5I
FmuI
<CjeI
>BbvI
<HphI
HpyCH4III
BslI
AAAAACAATACATACTGACAATGGCAGCAATTTCACCGGTGCTACGGTTAGGGCCGCCTG 4620
TTTTTGTTATGTATGACTGTTACCGTCGTTAAAGTGGCCACGATGCCAATCCCGGCGGAC
4574 4584 4593 4604 4617
4584 4593 4611
4584 4594 4611
4584 4595 4611
4589 4612
4595 4612
4595 4613
4596 4613
4613
4614
4614
TfiI
HinfI
Tsp509I
<Sth132I
Tsp509I
<FauI
ApoI
EcoRI
<AciI
>Tth111II
ApoI
>CstMI
TTGGTGGGCGGGAATCAAGCAGGAATTTGGAATTCCCTACAATCCCCAAAGTCAAGGAGT 4680
AACCACCCGCCCTTAGTTCGTCCTTAAACCTTAAGGGATGTTAGGGGTTTCAGTTCCTCA
4628 4636 4650 4674
4628 4643 4650
4629 4644
4632 4651
4632

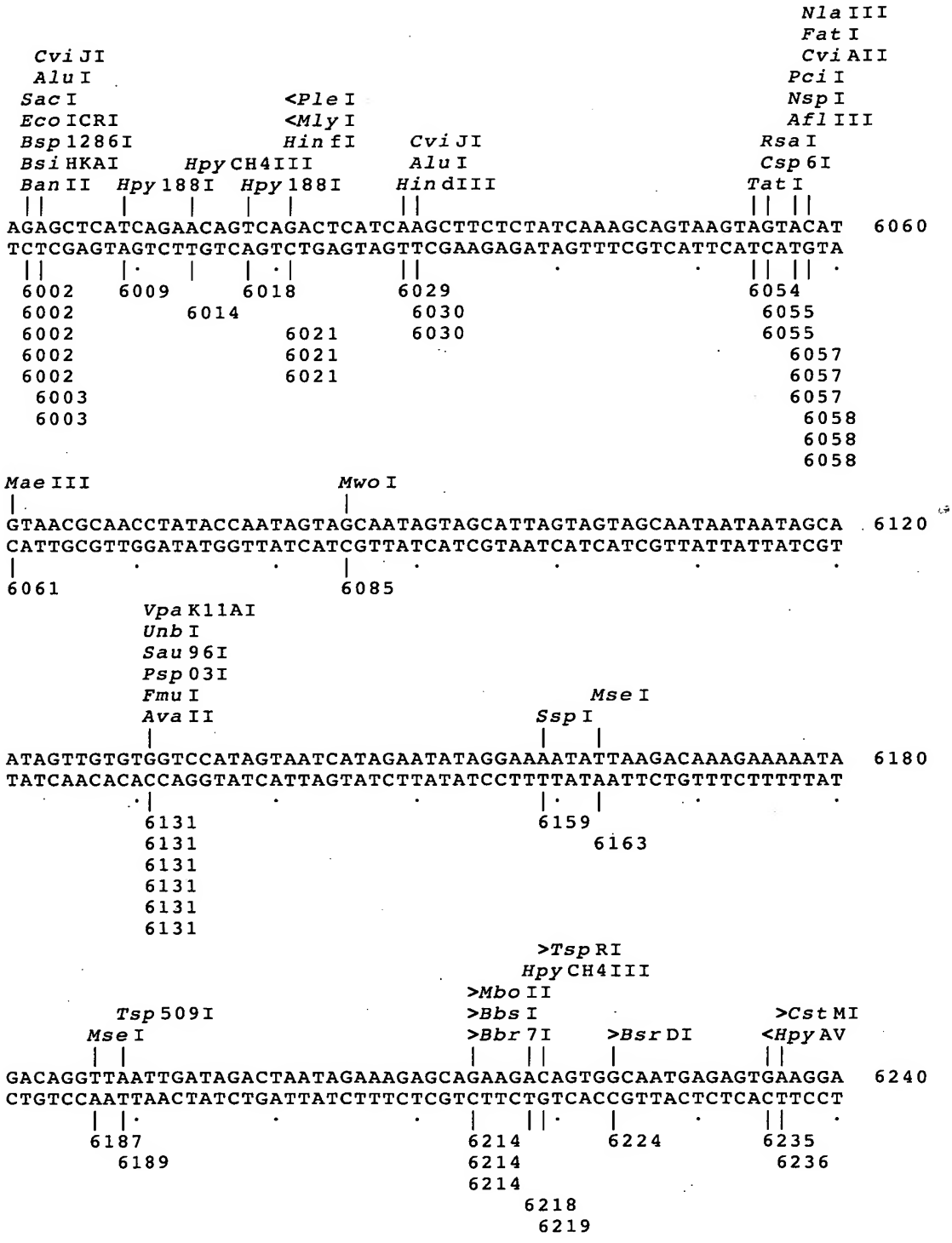

Cvi JI
Mbo I
Dpn I
Cha I
Bst KTI
Tfi I
Hin fI
>Tsp DTI
Mse I
Tsp 509I
Tsp 509I
AGTAGAATCTATGAATAAAGAATTAAAGAAAATTATAGGACAGGTAAGAGATCAGGCTGA 4740
TCATCTTAGATACTTATTTCTTAATTTCTTTTAATATCCTGTCCATTCTCTAGTCCGACT
4685
4685 4691
4701
4703
4711
4730
4730
4730
4730
4735
<Sts I
<Fok I
<Bst F5I
Mse I
Sml I
Afl II
Rsa I
Csp 6I
Tat I
<Tsp DTI
Dra I
Tsp 509I
cPPT
ACATCTTAAGACAGCAGTACAAATGGCAGTATTCATCCACAATTTTAAAAGAAAAGGGGG 4800
TGTAGAATTCTGTCGTCATGTTTACCGTCATAAGTAGGTGTTAAAATTTTCTTTTCCCCC
4745
4745
4746
4756
4757
4757
4772
4774
4774
4774
4781
4784
4785
>Bsg I
>Tsp RI
Hpy CH4III
Rsa I
Csp 6I
Hpy CH4V
Hpy 8I
Acc I
GATTGGGGGGTACAGTGCAGGGGAAAGAATAGTAGACATAATAGCAACAGACATACAAAC 4860
CTAACCCCCCATGTCACGTCCCCTTTCTTATCATCTGTATTATCGTTGTCTGTATGTTTG
4810
4810
4812
4813
4815
4816
4832
4832
CTS
Apo I
Apo I
Tsp 509I
Tsp 509I
Tsp 509I
Tsp 509I
<Sth 132I
>Bsm FI
TAAAGAATTACAAAAACAAATTACAAAAATTCAAAATTTTCGGGTTTATTACAGGGACAG 4920
ATTTCTTAATGTTTTTGTTTAATGTTTTTAAGTTTTAAAAGCCCAAATAATGTCCCTGTC
4866
4879
4887
4888
4894
4895
4901
4914
>Cje PI
Vpa K11AI
Unb I
Sau 96I
Psp 03I
Fmu I
Ava II
>Mnl I
<Bse RI
Cvi JI
Alu I
>Ssp D5I
>Hph I
Hpy 188III
<Hpy AV
CAGAAATCCACTTTGGAAAGGACCAGCAAAGCTCCTCTGGAAAGGTGAAGGGGCAGTAGT 4980
GTCTTTAGGTGAAACCTTTCCTGGTCGTTTCGAGGAGACCTTTCCACTTCCCCGTCATCA
4940
4940
4940
4940
4940
4940
4943
4950
4950
4952
4954
4956
4964
4964
4967

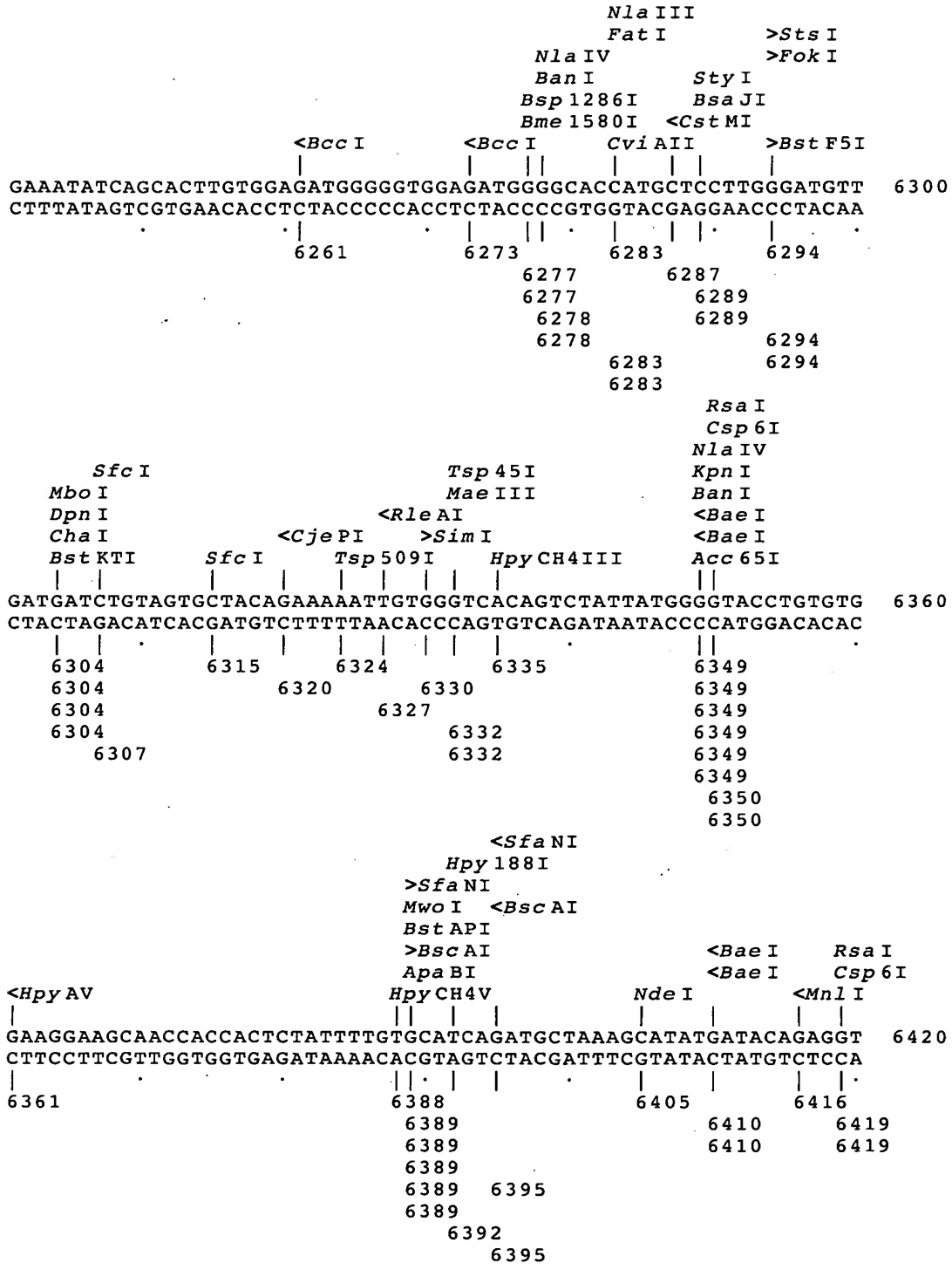

Mbo I
Dpn I
Cha I
Bst KTI
Tsp 45I
Mae III
>Mbo II
AATACAAGATAATAGTGACATAAAAGTAGTGCCAAGAAGAAAAGCAAAGATCATTAGGGA 5040
TTATGTTCTATTATCACTGTATTTTCATCACGGTTCTTCTTTTCGTTTCTAGTAATCCCT
4995 5016 5029
4995 5029
5029
5029
>Ssp D5I
<Mnl I
<Bsp MI
>Sts I
<Bcc I >Hph I
Hpy 8I >Fok I
<Cje PI <AarI
Acc I >Bst F5I
TTATGGAAAACAGATGGCAGGTGATGATTGTGTGGCAAGTAGACAGGATGAGGATTAGAA 5100
AATACCTTTTGTCTACCGTCCACTACTAACACACCGTTCATCTGTCCTACTCCTAATCTT
5046 5057 5079 5086
5053 5060 5079 5086
5057 5086
5060 5090
>Sts I
Nla III
Bfa I <Bcc I
Fat I Nde I Cvi JI >Fok I
Cvi AII Msl I Alu I >Bst F5I
CATGGAAAAGTTTAGTAAAACACCATATGTATGTTTCAGGGAAAGCTAGGGGATGGTTTT 5160
GTACCTTTTCAAATCATTTTGTGGTATACATACAAAGTCCCTTTCGATCCCCTACCAAAA
5101 5124 5144 5151
5101 5124 5144 5151
5101 5146 5152
5151
<Sts I
<Sts I
Rsa I
<Fok I
Hpy 8I
>Mnl I
Csp 6I Bsl I
Cvi JI Xmn I Tat I <Fok I Bfa I
Msl I >Tsp DTI <Bst F5I Hpy 188I <Bst F5I
ATAGACATCACTATGAAAGCCCTCATCCAAGAATAAGTTCAGAAGTACACATCCCACTAG 5220
TATCTGTAGTGATACTTTCGGGAGTAGGTTCTTATTCAAGTCTTCATGTGTAGGGTGATC
5166 5173 5178 5184 5191 5199 5204 5210 5210 5217
5181 5205 5213
5184 5205
5184 5205
5210
Bfa I
<Sfa NI
<Bsc AI
>Sts I
>Bsr I
>Fok I
Hpy CH4V
<Bsm AI
>Bst F5I
>Sim I
<Cje PI
GGGATGCTAGATTGGTAATAACAACATATTGGGGTCTGCATACAGGAGAAAGAGACTGGC 5280
CCCTACGATCTAACCATTATTGTTGTATAACCCCAGACGTATGTCCTCTTTCTCTGACCG
5222 5252 5268
5222 5257 5272
5222 5275
5223
5223
5227

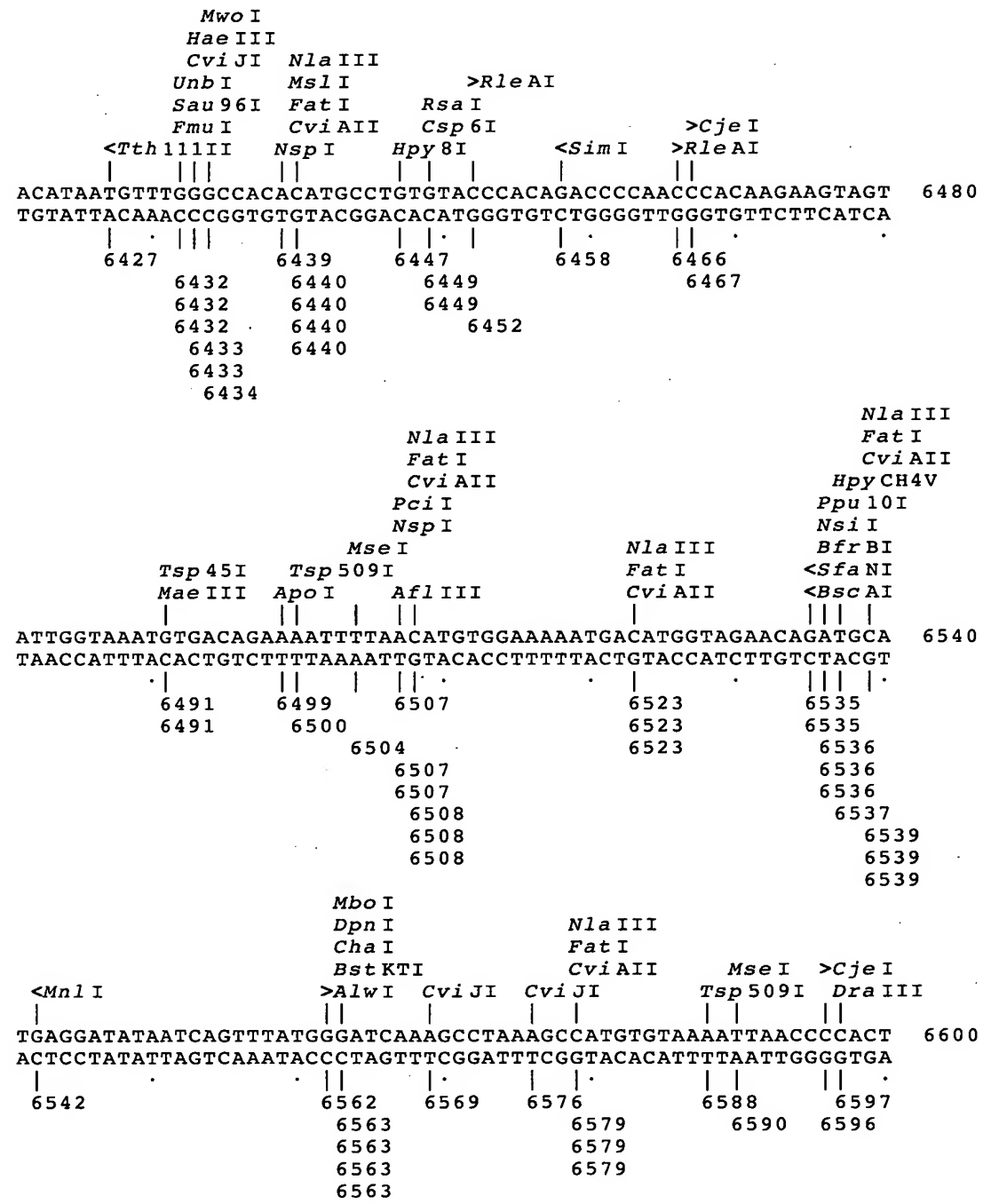

Pfl MI
Bsl I
>Bsm AI
>Ple I
>Mly I
>Sim I
Hin fI
<Mnl I
<Sim I
Hpy 8I
Acc I
ATTTGGGTCAGGGAGTCTCCATAGAATGGAGGAAAAAGAGATATAGCACACAAGTAGACC 5340
TAAACCCAGTCCCTCAGAGGTATCTTACCTCCTTTTTCTCTATATCGTGTGTTCATCTGG
5285
5293
5293
5293
5295
5299
5299
5309
5334
5334
5337
<Ple I
<Mly I
Alw NI
Hpy 188I
<Tsp DTI
Bfa I
Tsp 509I
Hpy CH4III
Hin fI
CTGAACTAGCAGACCAACTAATTCATCTGTATTACTTTGACTGTTTTTCAGACTCTGCTA 5400
GACTTGATCGTCTGGTTGATTAAGTAGACATAATGAAACTGACAAAAAGTCTGAGACGAT
5346
5360
5362
5380
5388
5389
5391
5391
Bfa I
Hae III
Sty I
Cvi JI
Bsa JI
Stu I
Avr II
Hae I
Cvi JI
>Tth 111II
TAAGAAAGGCCTTATTAGGACACATAGTTAGCCCTAGGTGTGAATATCAAGCAGGACATA 5460
ATTCTTTCCGGAATAATCCTGTGTATCAATCGGGATCCACACTTATAGTTCGTCCTGTAT
5407
5407
5408
5408
5430
5433
5433
5433
5434
5448
Mbo I
Dpn I
Cha I
Bst KTI
Tse I
Mse I
Fnu 4HI
Bth CI
Bst YI
>Alw I
<Cje I
Mwo I
Bfa I
Ase I
>Bbv I
ACAAGGTAGGATCTCTACAATACTTGGCACTAGCAGCATTAATAACACCAAAAAAGATAA 5520
TGTTCCATCCTAGAGATGTTATGAACCGTGATCGTCGTAATTATTGTGGTTTTTTCTATT
5469
5469
5470
5470
5470
5470
5477
5487
5490
5493
5493
5493
5493
5498
5499

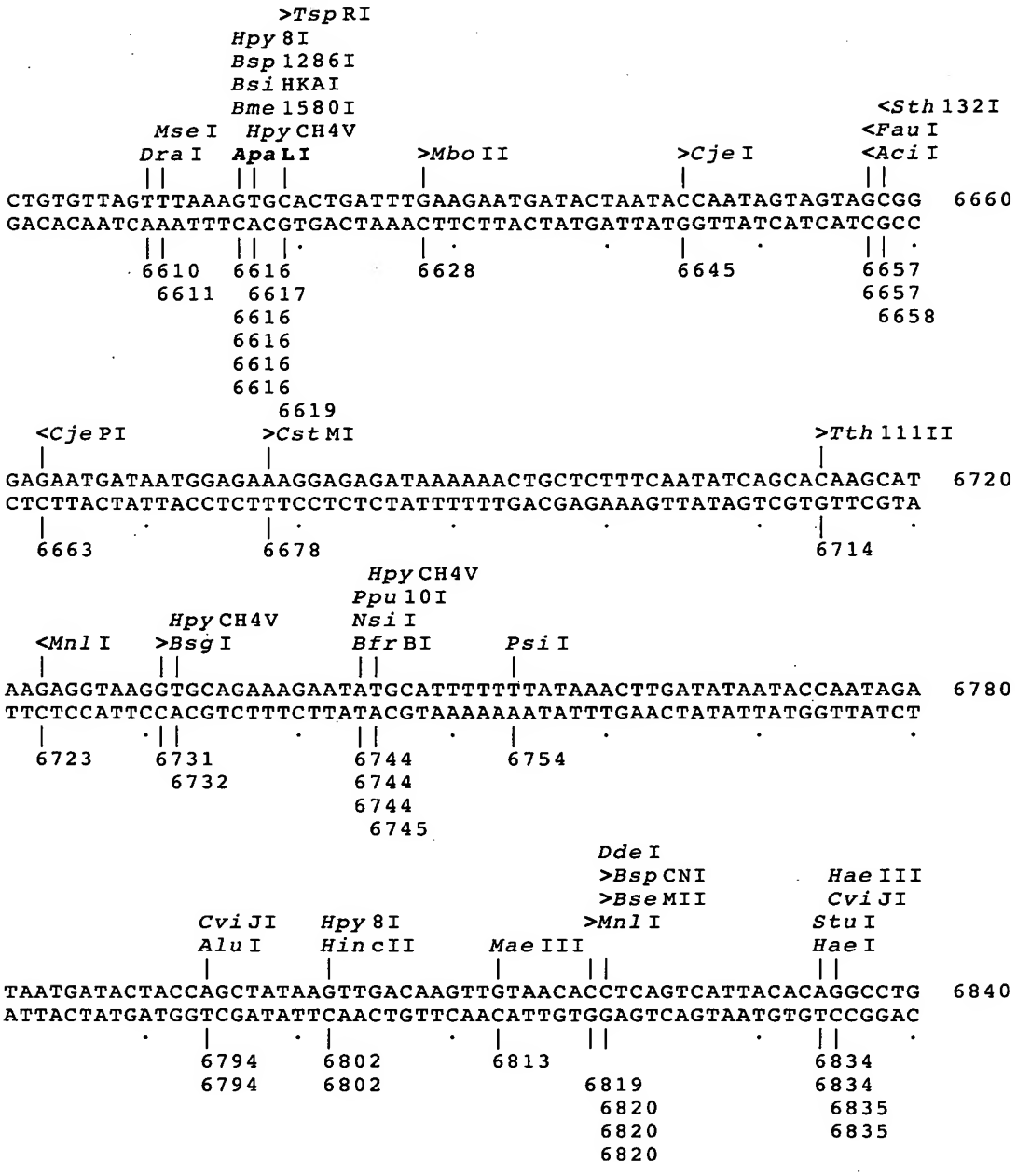

Mwo I
Cvi JI
Mae III
Bfa I
<Mnl I
<Cje PI
<Bcc I
Cvi JI
>Mbo II
>Bbs I
>Bbr 7I
AGCCACCTTTGCCTAGTGTTACGAAACTGACAGAGGATAGATGGAACAAGCCCCAGAAGA 5580
TCGGTGGAAACGGATCACAATGCTTTGACTGTCTCCTATCTACCTTGTTCGGGGTCTTCT
5521 5533 5553 5569 5576
5522 5538 5553 5560 5576
5576
Hae III
Cvi JI
Unb I
Sau 96I
Fmu I
Cvi JI
Nla IV
<Cje I
Sty I
Bsa JI
<Mnl I
Xcm I
>Tsp DTI
Cvi JI
Alu I
Bfa I
Cvi JI
Alu I
<Mnl I
Mse I
>Bse RI
Sml I
Afl II
Cvi JI
Alu I
CCAAGGGCCACAGAGGGAGCCACACAATGAATGGACACTAGAGCTTTTAGAGGAGCTTAA 5640
GGTTCCCGGTGTCTCCCTCGGTGTGTTACTTACCTGTGATCTCGAAAATCTCCTCGAATT
5581 5593 5600 5607 5618 5630
5581 5596 5604 5622 5630 5637
5585 5598 5622 5634
5585 5634
5585 5636
5586 5636
5586
Nla III
Fat I
Cvi AII
Sty I
Nco I
Btg I
Dde I
Bfa I
Sty I
Bsa JI
Cvi JI
Alu I
Bsa JI
Avr II
Nla IV
Cvi JI
<Bpu 10I
Cvi JI
Msl I
>Tsp DTI
>Tsp DTI
GAATGAAGCTGTTAGACATTTTCCTAGGATTTGGCTCCATGGCTTAGGGCAACATATCTA 5700
CTTACTTCGACAATCTGTAAAAGGATCCTAAACCGAGGTACCGAATCCCGTTGTATAGAT
5643 5663 5673 5681 5693 5700
5647 5663 5673 5682
5647 5663 5677
5664 5677 5683
5677
5677
5678
5678
5678
Mwo I
Bst API
Tsp 509I
Eco RI
Apa BI
Apo I
Hpy CH4V
<Bci VI
Cvi JI
TGAAACTTATGGGGATACTTGGGCAGGAGTGGAAGCCATAATAAGAATTCTGCAACAACT 5760
ACTTTGAATACCCCTATGAACCCGTCCTCACCTTCGGTATTATTCTTAAGACGTTGTTGA
5713 5734 5745 5751
5745 5752
5746
5752
5752

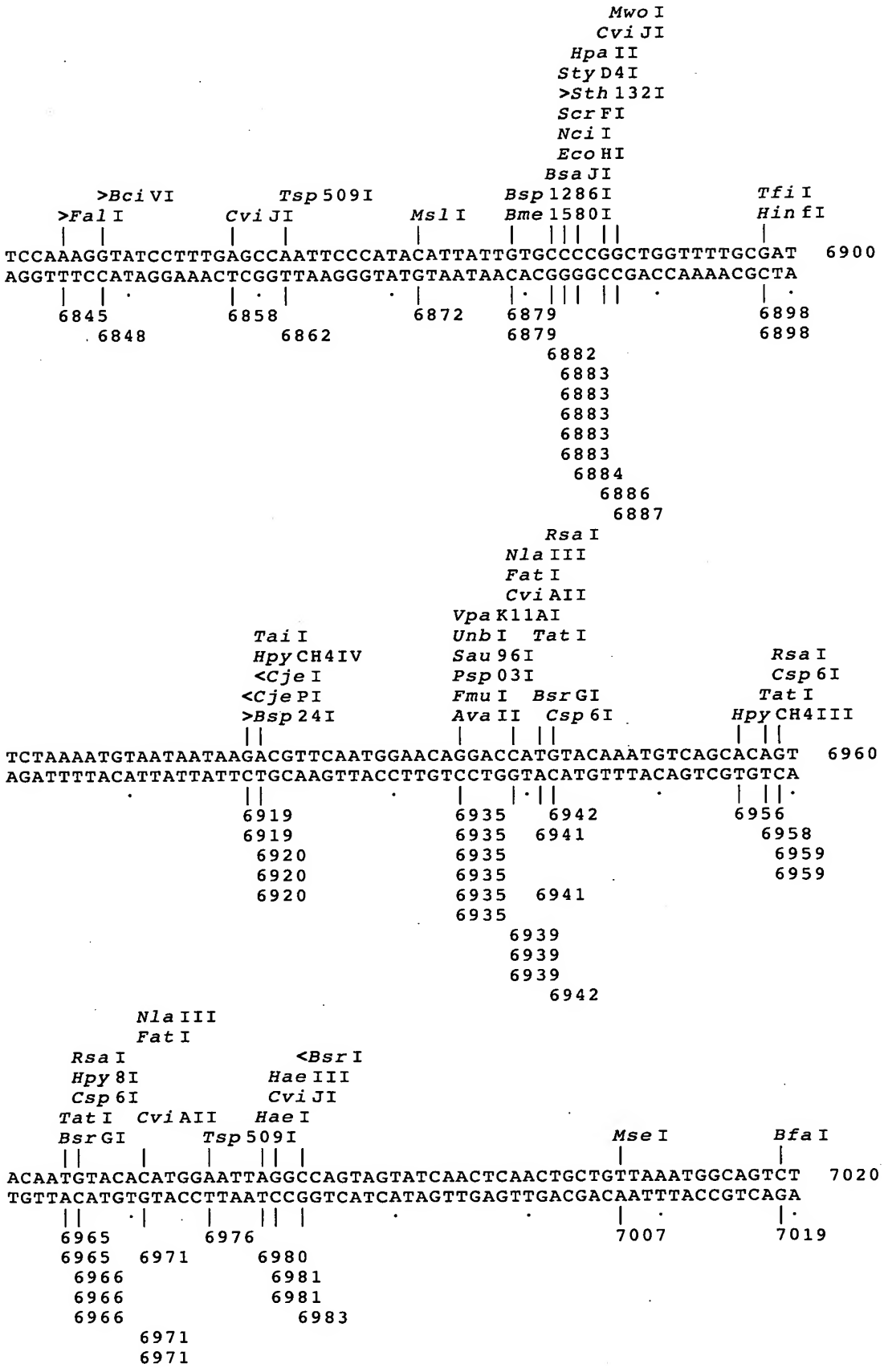

TaqI
SalI
Hpy8I
HincII
TspEsaBC3I
Hpy188I AccI MwoI MaeIII TaqI <MnlI
EsaBC3I
>BseRI
GCTGTTTATCCATTTTCAGAATTGGGTGTCGACATAGCAGAATAGGCGTTACTCGACAGA 5820
CGACAAATAGGTAAAAGTCTTAACCCACAGCTGTATCGTCTTATCCGCAATGAGCTGTCT
5776 5788 5797 5808 5819
5780 5789 5813
5788 5813 5819
5788
5788
5789
StyD4I
ScrFI
PspGI
BstNI
PfoI
StyD4I
ScrFI
PspGI
BstNI
MboI
DpnI
ChaI
BstKTI
MwoI
CviJI
<StsI
<FokI
<BstF5I
<BsrI BfaI Bsp1286I
CviJI <AlwI Ban II >SfaNI
NlaIV BstYI BfaI BsaJI >BscAI
GGAGAGCAAGAAATGGAGCCAGTAGATCCTAGACTAGAGCCCTGGAAGCATCCAGGAAGT 5880
CCTCTCGTTCTTTACCTCGGTCATCTAGGATCTGATCTCGGGACCTTCGTAGGTCCTTCA
5835 5844 5854 5860 5868
5837 5845 5857 5868
5839 5849 5857 5869
5845 5858 5869
5845 5859 5869
5845 5861 5871
5845 5861 5872
5861 5872
5861 5872
5872
TspMfeI 509I
RsaI
Csp6I
<BsrDI
<TspDTI
CviJI <Tth111II >FalI MwoI
CAGCCTAAAACTGCTTGTACCAATTGCTATTGTAAAAAGTGTTGCTTTCATTGCCAAGTT 5940
GTCGGATTTTGACGAACATGGTTAACGATAACATTTTTCACAACGAAAGTAACGGTTCAA
5882 5892 5917 5924
5897 5927
5897 5929
5901
5902
>SfaNI
<SapI
DdeI EcoNI MwoI >MboII
Bsu36I Bsl I <AciI Hpy99I
<TspDTI CviJI >BscAI >MboII <BsmAI <EarI
TGTTTCATAACAAAAGCCTTAGGCATCTCCTATGGCAGGAAGAAGCGGAGACAGCGACGA 6000
ACAAAGTATTGTTTTCGGAATCCGTAGAGGATACCGTCCTTCTTCGCCTCTGTCGCTGCT
5944 5955 5963 5979 5988 5999
5957 5969 5985 5995
5958 5969 5985 5999
5963 5999


NlaIII
FatI
CviAII
PciI
NspI
AflIII
CviJI
AluI
SacI
<PleI
EcoICRI
<MlyI
Bsp1286I
HinfI
CviJI
RsaI
BsiHKAI
HpyCH4III
AluI
CspI
BanII
Hpy188I
Hpy188I
HindIII
TatI
AGAGCTCATCAGAACAGTCAGACTCATCAAGCTTCTCTATCAAAGCAGTAAGTAGTACAT 6060
TCTCGAGTAGTCTTGTCAGTCTGAGTAGTTCGAAGAGATAGTTTCGTCATTCATCATGTA
6002 6009 6018 6029 6054
6002 6014 6030 6055
6002 6021 6030 6055
6002 6021 6057
6002 6021 6057
6003 6057
6003 6058
6058
6058
MaeIII
MwoI
GTAACGCAACCTATACCAATAGTAGCAATAGTAGCATTAGTAGTAGCAATAATAATAGCA 6120
CATTGCGTTGGATATGGTTATCATCGTTATCATCGTAATCATCATCGTTATTATTATCGT
6061 6085
VpaK11AI
UnbI
Sau96I
Psp03I
FmuI
MseI
AvaII
SspI
ATAGTTGTGTGGTCCATAGTAATCATAGAATATAGGAAAATATTAAGACAAAGAAAAATA 6180
TATCAACACACCAGGTATCATTAGTATCTTATATCCTTTTATAATTCTGTTTCTTTTTAT
6131 6159
6131 6163
6131
6131
6131
6131
>TspRI
HpyCH4III
>MboII
Tsp509I
>BbsI
>CstMI
MseI
>Bbr7I
>BsrDI
<HpyAV
GACAGGTTAATTGATAGACTAATAGAAAGAGCAGAAGACAGTGGCAATGAGAGTGAAGGA 6240
CTGTCCAATTAACTATCTGATTATCTTTCTCGTCTTCTGTCACCGTTACTCTCACTTCCT
6187 6214 6224 6235
6189 6214 6236
6214
6218
6219

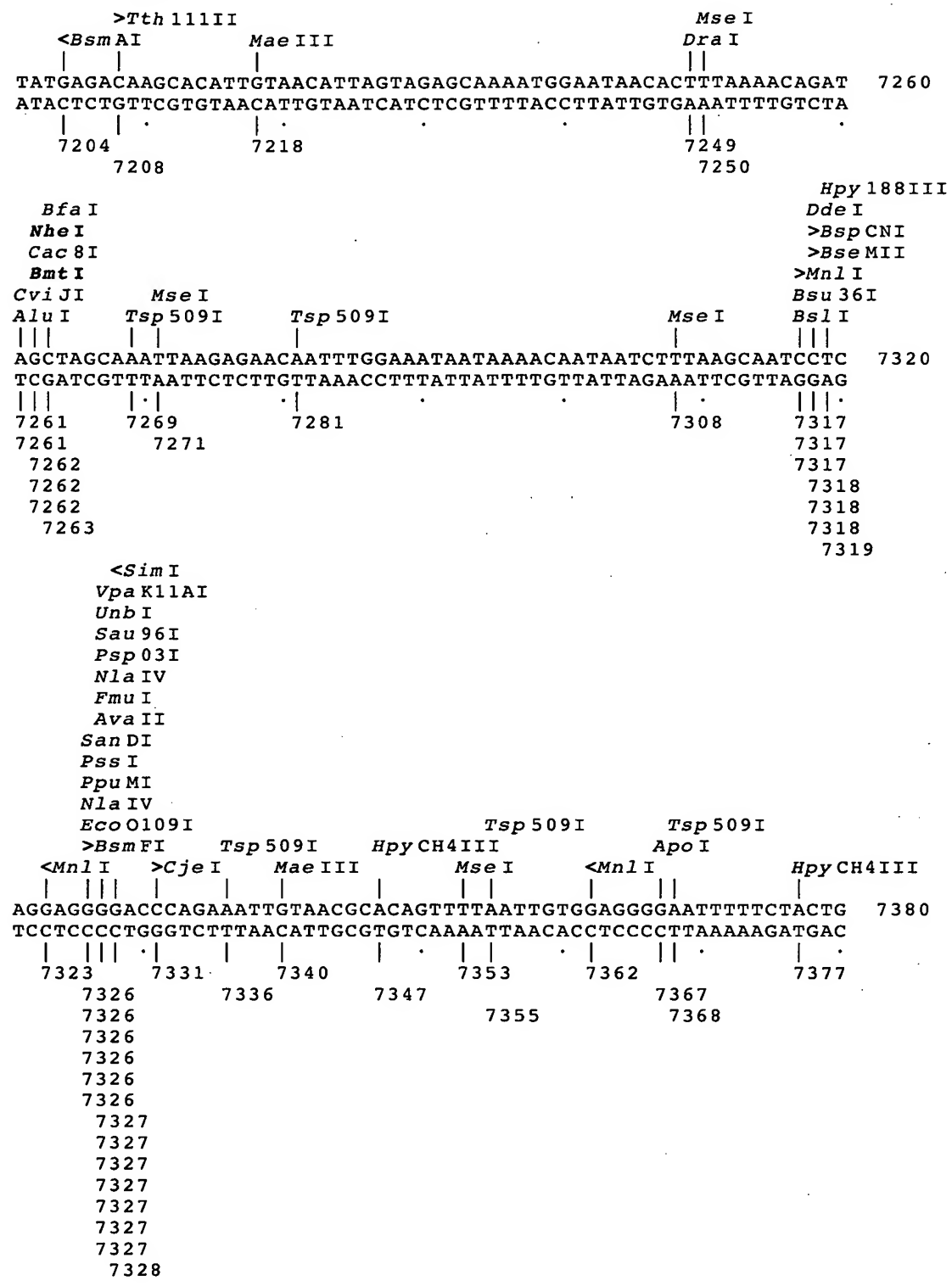

Nla III
Fat I
>Sts I
Nla IV
>Fok I
Ban I
Sty I
Bsp 1286I
Bsa JI
Bme 1580I
<Cst MI
<Bcc I
<Bcc I
Cvi AII
>Bst F5I
GAAATATCAGCACTTGTGGAGATGGGGGTGGAGATGGGGCACCATGCTCCTTGGGATGTT 6300
CTTTATAGTCGTGAACACCTCTACCCCCACCTCTACCCCGTGGTACGAGGAACCCTACAA
6261
6273
6283
6294
6277
6287
6277
6289
6278
6289
6278
6294
6283
6294
6283
Rsa I
Csp 6I
Nla IV
Sfc I
Tsp 45I
Kpn I
Mbo I
Mae III
Ban I
Dpn I
<Rle AI
<Bae I
Cha I
<Cje PI
>Sim I
<Bae I
Bst KTI
Sfc I
Tsp 509I
Hpy CH4III
Acc 65I
GATGATCTGTAGTGCTACAGAAAAATTGTGGGTCACAGTCTATTATGGGGTACCTGTGTG 6360
CTACTAGACATCACGATGTCTTTTTAACACCCAGTGTCAGATAATACCCCATGGACACAC
6304
6315
6324
6335
6349
6304
6320
6330
6349
6304
6327
6349
6304
6332
6349
6307
6332
6349
6349
6350
6350
<Sfa NI
Hpy 188I
>Sfa NI
Mwo I
<Bsc AI
Bst API
>Bsc AI
<Bae I
Rsa I
Apa BI
<Bae I
Csp 6I
<Hpy AV
Hpy CH4V
Nde I
<Mnl I
GAAGGAAGCAACCACCACTCTATTTTGTGCATCAGATGCTAAAGCATATGATACAGAGGT 6420
CTTCCTTCGTTGGTGGTGAGATAAAACACGTAGTCTACGATTTCGTATACTATGTCTCCA
6361
6388
6405
6416
6389
6410
6419
6389
6410
6419
6389
6389
6395
6389
6392
6395

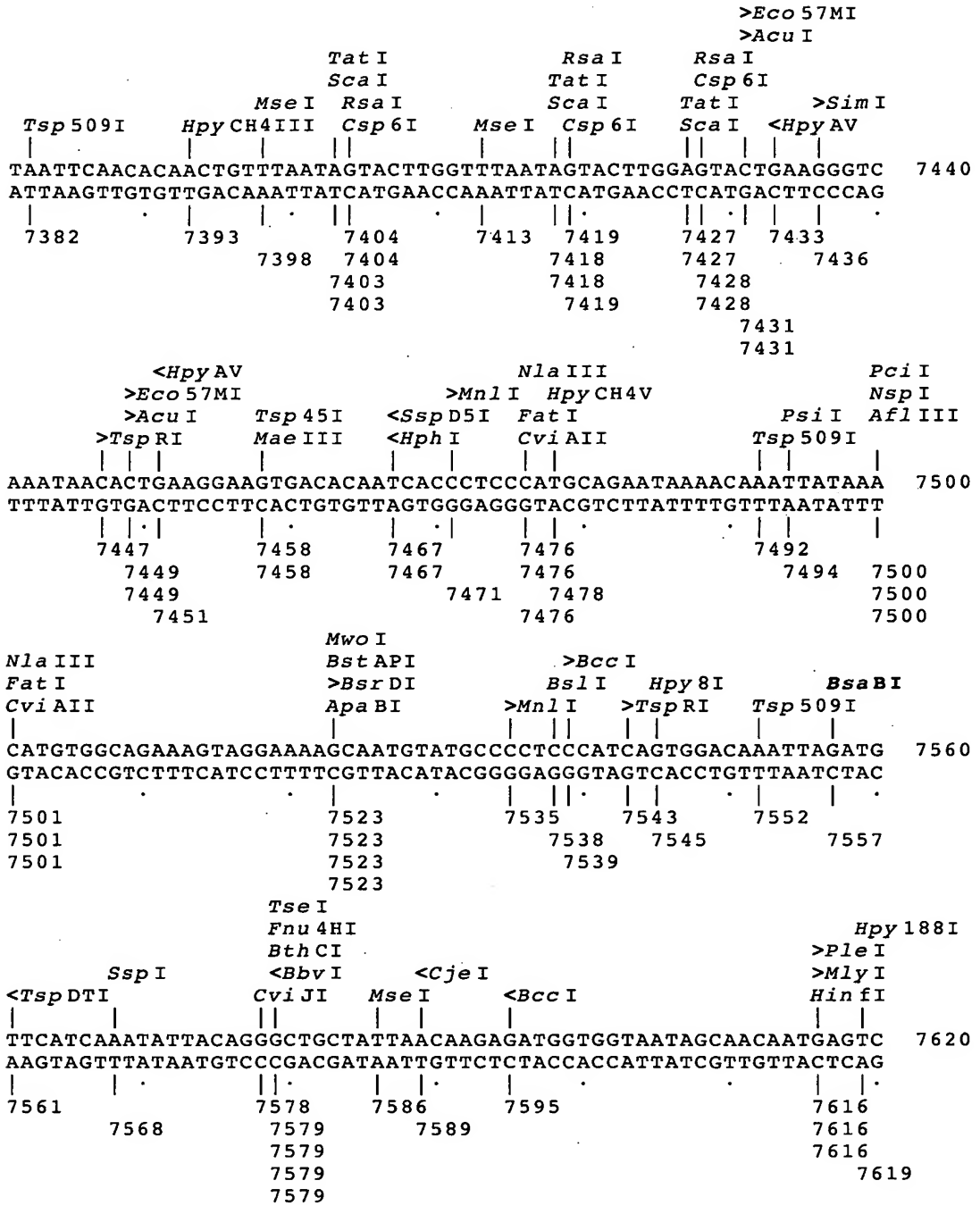

Mwo I
Hae III
Cvi JI
Nla III
Unb I
Msl I
>Rle AI
Sau 96I
Fat I
Rsa I
Fmu I
Cvi AII
Csp 6I
>Cje I
<Tth 111II
Nsp I
Hpy 8I
<Sim I
>Rle AI
ACATAATGTTTGGGCCACACATGCCTGTGTACCCACAGACCCCAACCCACAAGAAGTAGT 6480
TGTATTACAAACCCGGTGTGTACGGACACATGGGTGTCTGGGGTTGGGTGTTCTTCATCA
6427
6439
6447
6458
6466
6432
6440
6449
6467
6432
6440
6449
6432
6440
6452
6433
6440
6433
6434
Nla III
Fat I
Cvi AII
Hpy CH4V
Ppu 10I
Nsi I
Bfr BI
<Sfa NI
<Bsc AI
Nla III
Fat I
Cvi AII
Pci I
Nsp I
Mse I
Tsp 45I
Mae III
Tsp 509I
Apo I
Afl III
Nla III
Fat I
Cvi AII
ATTGGTAAATGTGACAGAAAATTTTAACATGTGGAAAAATGACATGGTAGAACAGATGCA 6540
TAACCATTTACACTGTCTTTTAAAATTGTACACCTTTTTACTGTACCATCTTGTCTACGT
6491
6499
6507
6523
6535
6491
6500
6523
6535
6504
6523
6536
6507
6536
6507
6536
6508
6537
6508
6539
6508
6539
6539
Mbo I
Dpn I
Cha I
Bst KTI
Nla III
Fat I
Cvi AII
Mse I
>Cje I
<Mnl I
>Alw I
Cvi JI
Cvi JI
Tsp 509I
Dra III
TGAGGATATAATCAGTTTATGGGATCAAAGCCTAAAGCCATGTGTAAAATTAACCCCACT 6600
ACTCCTATATTAGTCAAATACCCTAGTTTCGGATTTCGGTACACATTTTAATTGGGGTGA
6542
6562
6569
6576
6588
6597
6563
6579
6590
6596
6563
6579
6563
6579
6563

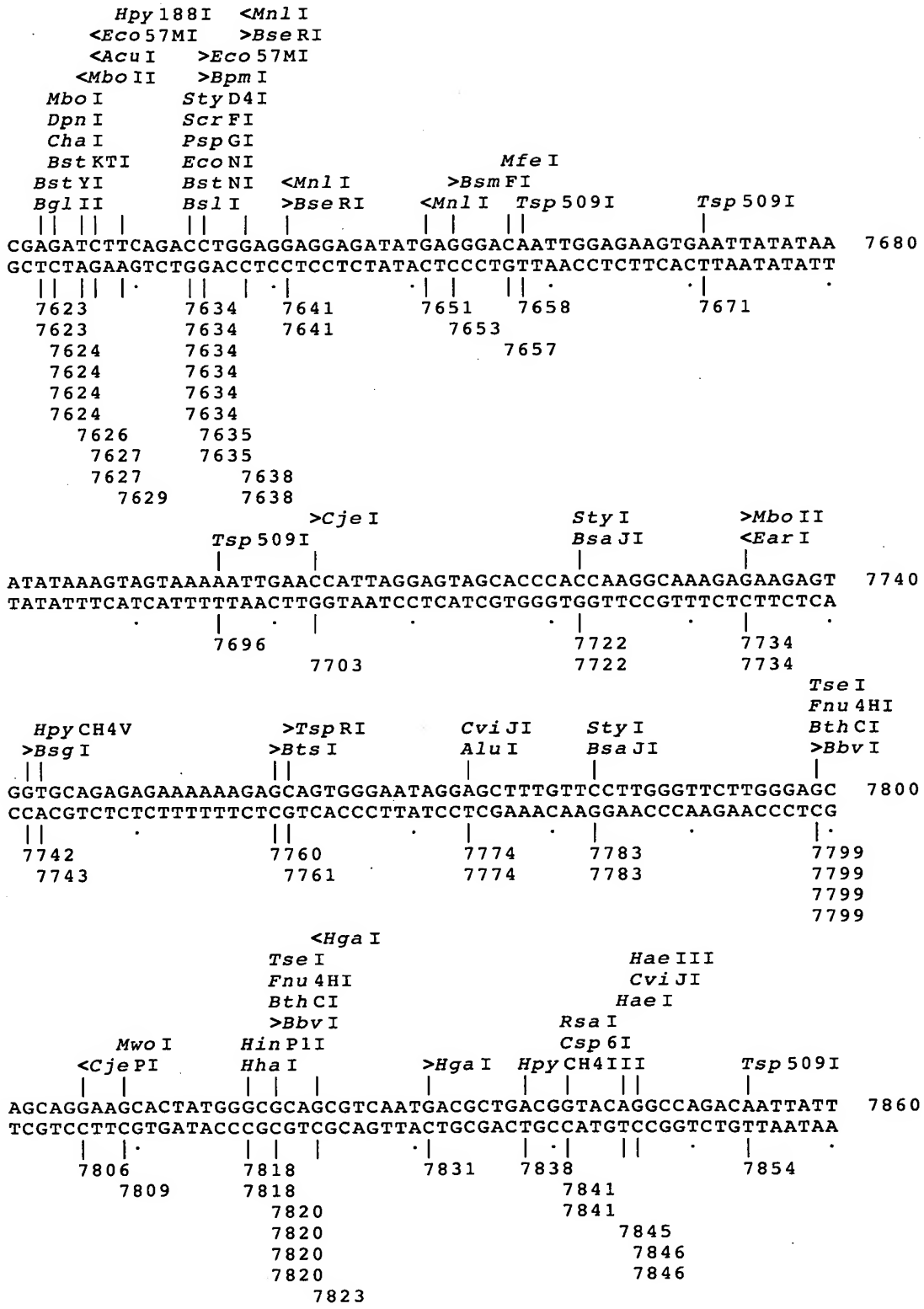

>TspRI
HpyBsp1286IBsiHKAIBme1580I
<Sth132I
MseI HpyCH4V
<FauI
DraI ApaLI >MboII >CjeI <AciI
CTGTGTTAGTTTAAAGTGCACTGATTTGAAGAATGATACTAATACCAATAGTAGTAGCGG 6660
GACACAATCAAATTTCACGTGACTAAACTTCTTACTATGATTATGGTTATCATCATCGCC
6610 6616 6628 6645 6657
6611 6617 6657
6616 6658
6616
6616
6616
6619
<CjePI >CstMI >Tth111II
GAGAATGATAATGGAGAAAGGAGAGATAAAAAACTGCTCTTTCAATATCAGCACAAGCAT 6720
CTCTTACTATTACCTCTTTCCTCTCTATTTTTTGACGAGAAAGTTATAGTCGTGTTCGTA
6663 6678 6714
HpyCH4V
Ppu10I
HpyCH4V NsiI
<MnlI >BsgI BfrBI PsiI
AAGAGGTAAGGTGCAGAAAGAATATGCATTTTTTTATAAACTTGATATAATACCAATAGA 6780
TTCTCCATTCCACGTCTTTCTTATACGTAAAAAAATATTTGAACTATATTATGGTTATCT
6723 6731 6744 6754
6732 6744
6744
6745
DdeI
>BspCNI HaeIII
>BseMII CviJI
CviJI HpyBsp8I >MnlI StuI
AluI HincII MaeIII HaeI
TAATGATACTACCAGCTATAAGTTGACAAGTTGTAACACCTCAGTCATTACACAGGCCTG 6840
ATTACTATGATGGTCGATATTCAACTGTTCAACATTGTGGAGTCAGTAATGTGTCCGGAC
6794 6802 6813 6819 6834
6794 6802 6820 6834
6820 6835
6820 6835
6820

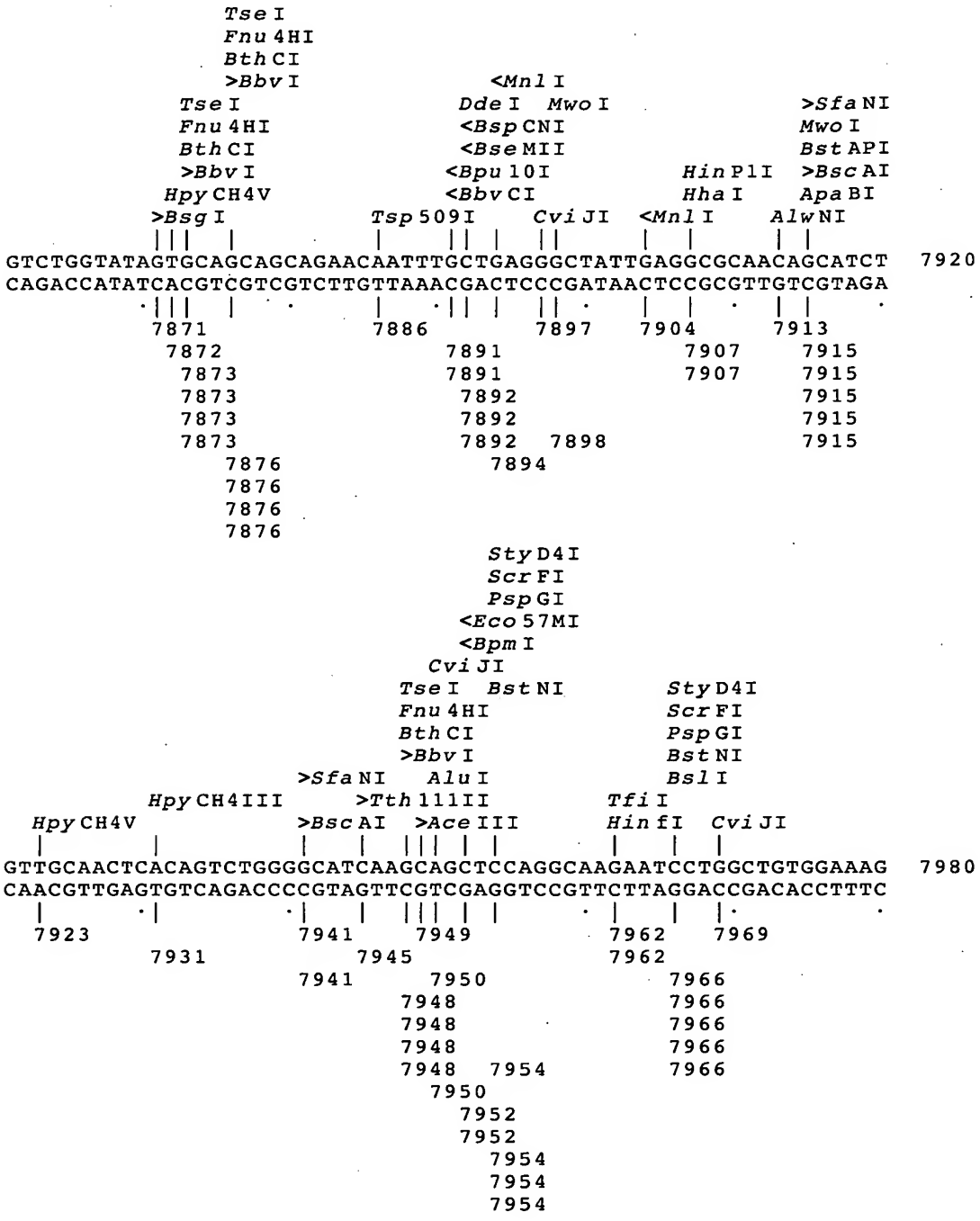

Mwo I
Cvi JI
Hpa II
Sty D4I
>Sth 132I
Scr FI
Nci I
Eco HI
Bsa JI
>Bci VI
Tsp 509I
Bsp 1286I
Tfi I
>Fal I
Cvi JI
Msl I
Bme 1580I
Hin fI
TCCAAAGGTATCCTTTGAGCCAATTCCCATACATTATTGTGCCCCGGCTGGTTTTGCGAT 6900
AGGTTTCCATAGGAAACTCGGTTAAGGGTATGTAATAACACGGGGCCGACCAAAACGCTA
6845
6858
6872
6879
6898
6848
6862
6879
6898
6882
6883
6883
6883
6883
6883
6884
6886
6887
Rsa I
Nla III
Fat I
Cvi AII
Vpa K11AI
Tai I
Unb I
Tat I
Hpy CH4IV
Sau 96I
Rsa I
<Cje I
Psp 03I
Csp 6I
<Cje PI
Fmu I
Bsr GI
Tat I
>Bsp 24I
Ava II
Csp 6I
Hpy CH4III
TCTAAAATGTAATAATAAGACGTTCAATGGAACAGGACCATGTACAAATGTCAGCACAGT 6960
AGATTTTACATTATTATTCTGCAAGTTACCTTGTCCTGGTACATGTTTACAGTCGTGTCA
6919
6935
6942
6956
6919
6935
6941
6958
6920
6935
6959
6920
6935
6959
6920
6935
6941
6935
6939
6939
6939
6942
Nla III
Fat I
Rsa I
<Bsr I
Hpy 8I
Hae III
Csp 6I
Cvi JI
Tat I
Cvi AII
Hae I
Bsr GI
Tsp 509I
Mse I
Bfa I
ACAATGTACACATGGAATTAGGCCAGTAGTATCAACTCAACTGCTGTTAAATGGCAGTCT 7020
TGTTACATGTGTACCTTAATCCGGTCATCATAGTTGAGTTGACGACAATTTACCGTCAGA
6965
6976
7007
7019
6965
6971
6980
6966
6981
6966
6981
6966
6983
6971
6971

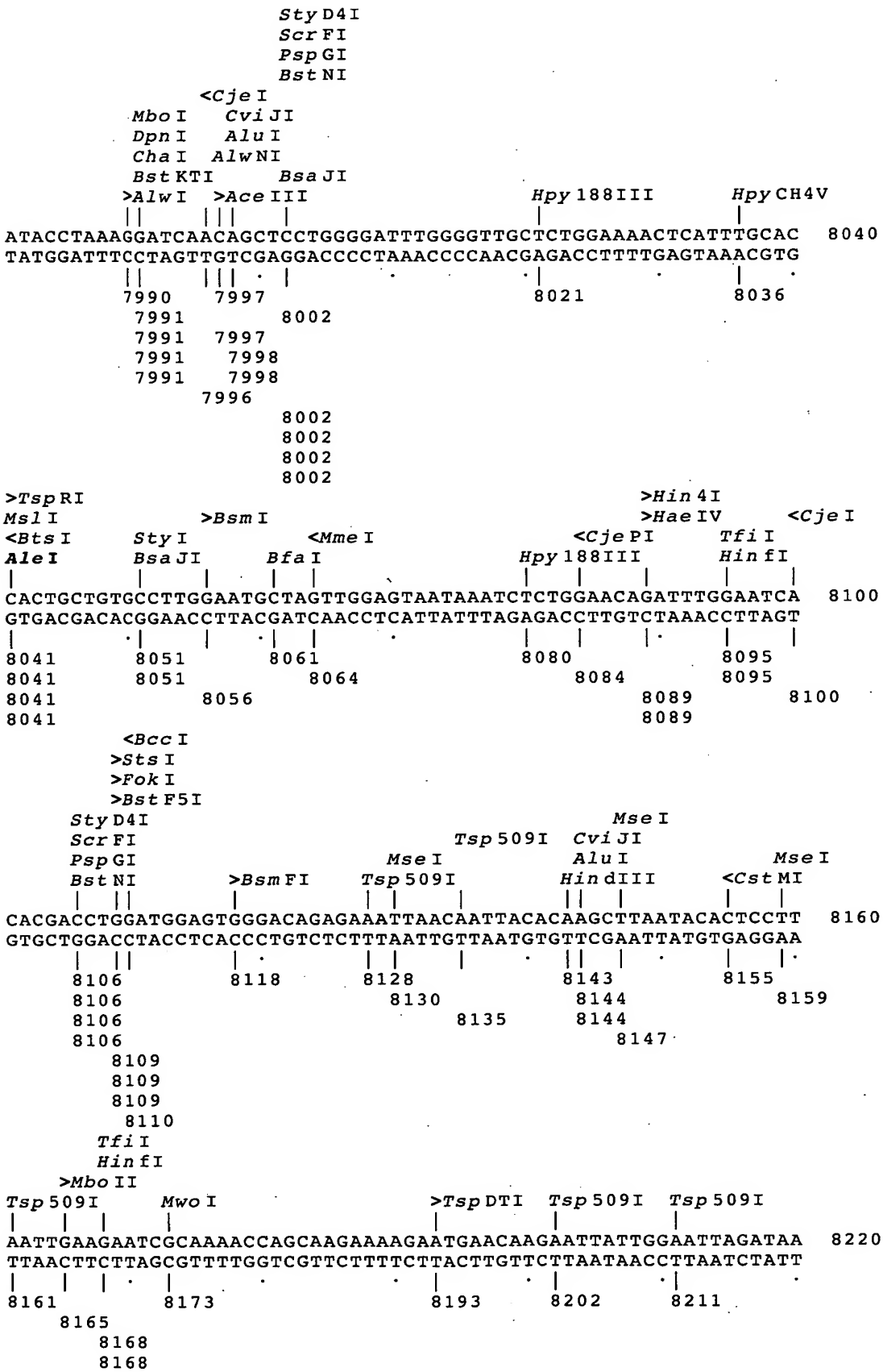
Mbo I
Dpn I
Cha I
<Mnl I
Bst KTI
>Mbo II
Bst YI
<Ear I
Bgl II
>Tsp GWI
>Mbo II
Tsp 509I
Tsp 509I
AGCAGAAGAAGAGGTAGTAATTAGATCTGTCAATTTCACGGACAATGCTAAAACCATAAT 7080
TCGTCTTCTTCTCCATCATTAATCTAGACAGTTAAAGTGCCTGTTACGATTTTGGTATTA
7025
7039
7052
7028
7043
7058
7028
7043
7031
7044
7044
7044
7044
Cvi JI
Alu I
Rsa I
Pvu II
Tat I
Msp A1I
Tsp 509I
Rsa I
Mse I
Csp 6I
Csp 6I
Ase I
Bsr GI
Tat I
Sfc I
Tsp 509I
<Sim I
AGTACAGCTGAACACATCTGTAGAAATTAATTGTACAAGACCCAACAACAATACAAGAAA 7140
TCATGTCGACTTGTGTAGACATCTTTAATTAACATGTTCTGGGTTGTTGTTATGTTCTTT
7081
7098
7105
7119
7082
7106
7112
7082
7107
7113
7085
7109
7085
7112
7086
7113
7086
Sty D4I
Scr FI
Psp GI
Bst NI
Vpa K11AI
Unb I
Sau 96I
Psp 03I
Fmu I
<Tsp GWI
Ava II
Tfi I
>Bci VI
<Mnl I
Hin fI
Hpy 188III
Bsa JI
Mae III
AAGAATCCGTATCCAGAGAGGACCAGGGAGAGCATTTGTTACAATAGGAAAAATAGGAAA 7200
TTCTTAGGCATAGGTCTCTCCTGGTCCCTCTCGTAAACAATGTTATCCTTTTTATCCTTT
7143
7152
7163
7178
7143
7149
7158
7146
7160
7160
7160
7160
7160
7160
7163
7163
7163
7163

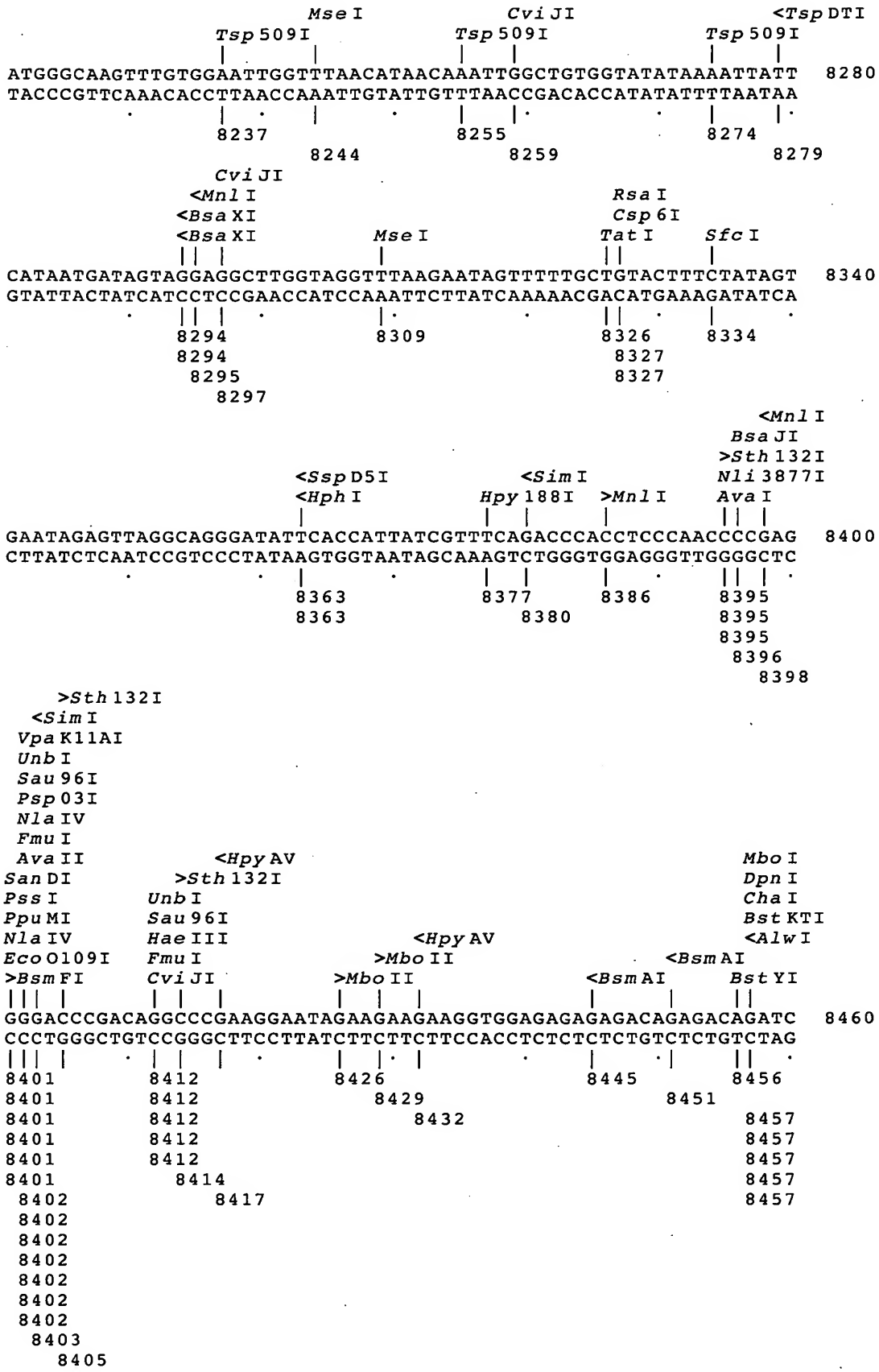

>Tth 111II
<BsmAI
MaeIII
MseI
DraI
TATGAGACAAGCACATTGTAACATTAGTAGAGCAAAATGGAATAACACTTTAAAACAGAT 7260
ATACTCTGTTCGTGTAACATTGTAATCATCTCGTTTTACCTTATTGTGAAATTTTGTCTA
7204
7208
7218
7249
7250
Hpy188III
BfaI
NheI
Cac8I
BmtI
DdeI
>BspCNI
>BseMII
>MnlI
Bsu36I
BslI
CviJI
AluI
MseI
Tsp509I
Tsp509I
MseI
AGCTAGCAAATTAAGAGAACAATTTGGAAATAATAAAACAATAATCTTTAAGCAATCCTC 7320
TCGATCGTTTAATTCTCTTGTTAAACCTTTATTATTTTGTTATTAGAAATTCGTTAGGAG
7261
7261
7262
7262
7262
7263
7269
7271
7281
7308
7317
7317
7317
7318
7318
7318
7319
<SimI
VpaK11AI
UnbI
Sau96I
Psp03I
NlaIV
FmuI
AvaII
SanDI
PssI
PpuMI
NlaIV
Eco0109I
>BsmFI
Tsp509I
HpyCH4III
Tsp509I
ApoI
Tsp509I
<MnlI
>CjeI
MaeIII
MseI
<MnlI
HpyCH4III
AGGAGGGGACCCAGAAATTGTAACGCACAGTTTTAATTGTGGAGGGGAATTTTTCTACTG 7380
TCCTCCCCTGGGTCTTTAACATTGCGTGTCAAAATTAACACCTCCCCTTAAAAAGATGAC
7323
7326
7326
7326
7326
7326
7326
7327
7327
7327
7327
7327
7327
7327
7328
7331
7336
7340
7347
7353
7355
7362
7367
7368
7377

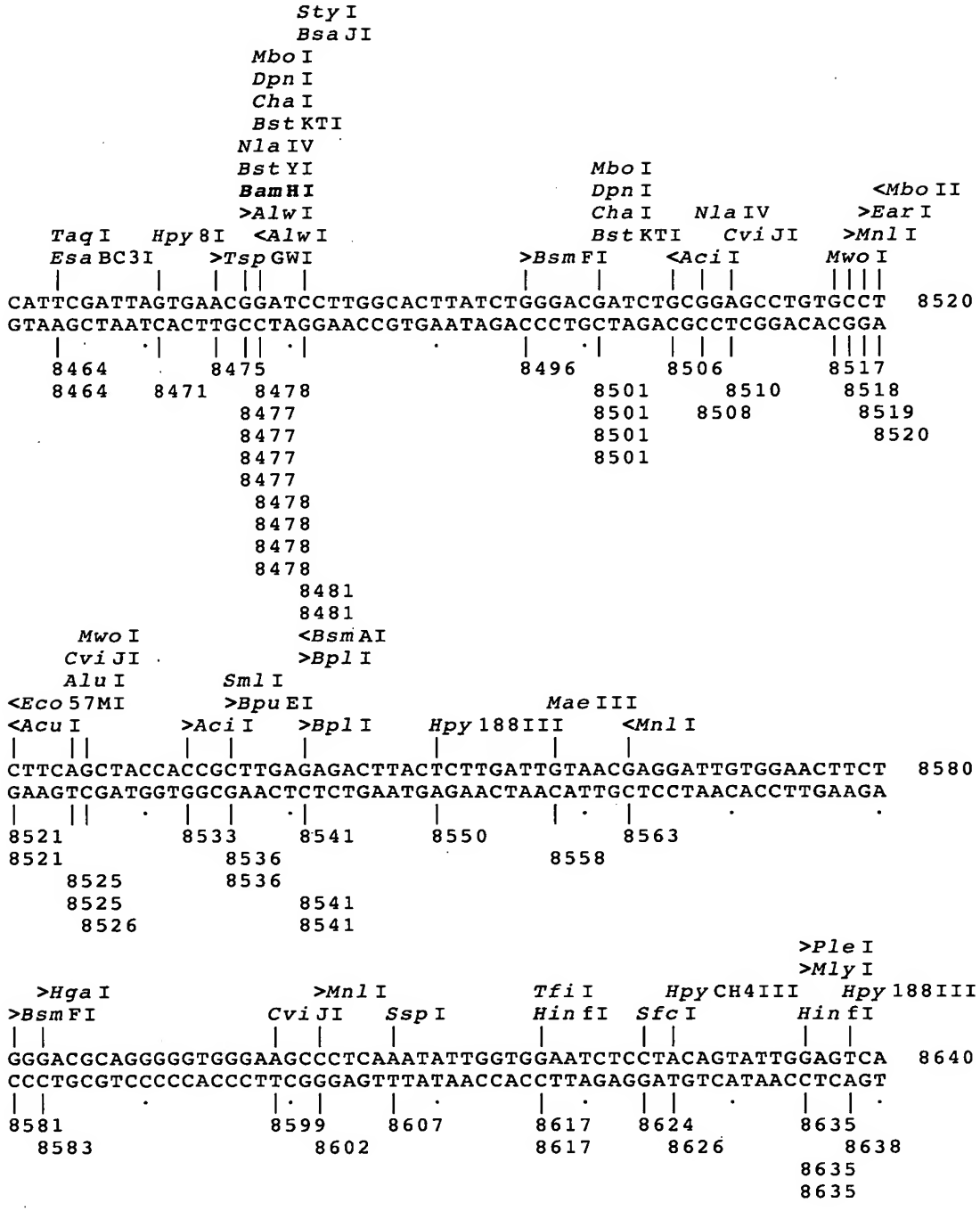

>Eco57MI
>AcuI
TatI
ScaI
RsaI
TatI
ScaI
RsaI
Csp6I
MseI
RsaI
ScaI
TatI
>SimI
Tsp509I
HpyCH4III
Csp6I
MseI
Csp6I
ScaI
<HpyAV
TAATTCAACACAACTGTTTAATAGTACTTGGTTTAATAGTACTTGGAGTACTGAAGGGTC 7440
ATTAAGTTGTGTTGACAAATTATCATGAACCAAATTATCATGAACCTCATGACTTCCCAG
7382 7393 7398 7404 7403 7403 7404 7413 7418 7419 7418 7427 7427 7428 7428 7431 7431 7433 7436
<HpyAV
>Eco57MI
>AcuI
>TspRI
Tsp45I
MaeIII
NlaIII
>MnlI
HpyCH4V
<SspD5I
<HphI
FatI
CviAII
PsiI
Tsp509I
PciI
NspI
AflIII
AAATAACACTGAAGGAAGTGACACAATCACCCTCCCATGCAGAATAAAACAAATTATAAA 7500
TTTATTGTGACTTCCTTCACTGTGTTAGTGGGAGGGTACGTCTTATTTTGTTTAATATTT
7447 7449 7449 7451 7458 7458 7467 7467 7471 7476 7476 7476 7478 7492 7494 7500 7500 7500
MwoI
NlaIII
FatI
CviAII
BstAPI
>BsrDI
ApaBI
>BccI
BslI
>MnlI
Hpy8I
>TspRI
Tsp509I
BsaBI
CATGTGGCAGAAAGTAGGAAAAGCAATGTATGCCCCTCCCATCAGTGGACAAATTAGATG 7560
GTACACCGTCTTTCATCCTTTTCGTTACATACGGGGAGGGTAGTCACCTGTTTAATCTAC
7501 7501 7501 7523 7523 7523 7523 7535 7538 7539 7543 7545 7552 7557
TseI
Fnu4HI
BthCI
<BbvI
SspI
<TspDTI
CviJI
MseI
<CjeI
<BccI
Hpy188I
>PleI
>MlyI
HinfI
TTCATCAAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTAATAGCAACAATGAGTC 7620
AAGTAGTTTATAATGTCCCGACGATAATTGTTCTCTACCACCATTATCGTTGTTACTCAG
7561 7568 7578 7579 7579 7579 7579 7586 7589 7595 7616 7616 7616 7619

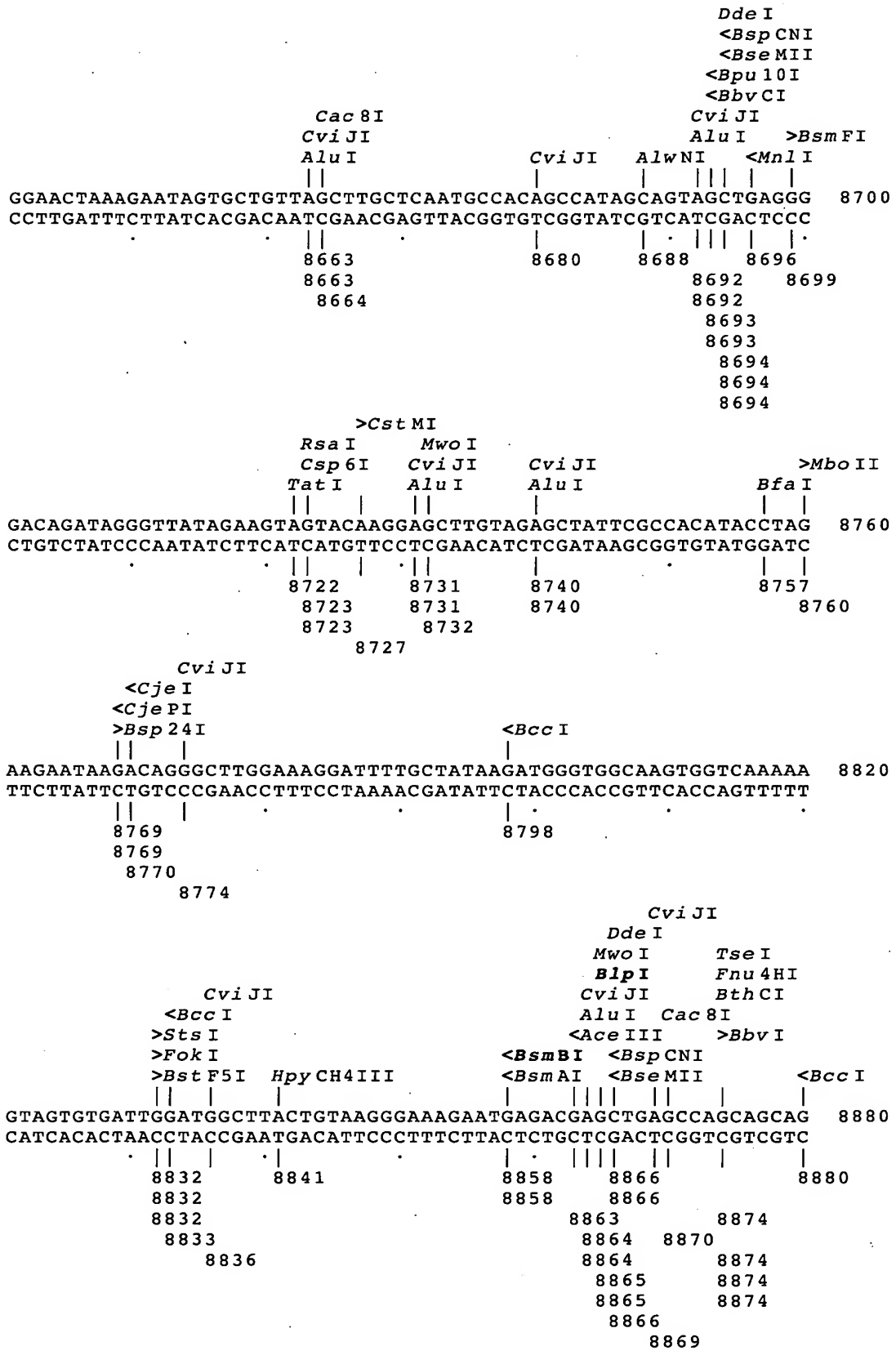

Hpy188I <Mnl I
<Eco57MI >BseRI
<AcuI >Eco57MI
<MboII >BpmI
MboI StyD4I
DpnI ScrFI
ChaI PspGI
BstKTI EcoNI MfeI
BstYI BstNI <MnlI >BsmFI
BglII BslI >BseRI <MnlI Tsp509I Tsp509I
CGAGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAA 7680
GCTCTAGAAGTCTGGACCTCCTCCTCTATACTCCCTGTTAACCTCTTCACTTAATATATT
7623 7634 7641 7651 7658 7671
7623 7634 7641 7653
7624 7634 7657
7624 7634
7624 7634
7624 7634
7626 7635
7627 7635
7627 7638
7629 7638
>CjeI StyI >MboII
Tsp509I BsaJI <EarI
ATATAAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAAGAGAAGAGT 7740
TATATTTCATCATTTTTAACTTGGTAATCCTCATCGTGGGTGGTTCCGTTTCTCTTCTCA
7696 7722 7734
7703 7722 7734
TseI
Fnu4HI
HpyCH4V >TspRI CviJI StyI BthCI
>BsgI >BtsI AluI BsaJI >BbvI
GGTGCAGAGAGAAAAAAGAGCAGTGGGAATAGGAGCTTTGTTCCTTGGGTTCTTGGGAGC 7800
CCACGTCTCTCTTTTTTCTCGTCACCCTTATCCTCGAAACAAGGAACCCAAGAACCCTCG
7742 7760 7774 7783 7799
7743 7761 7774 7783 7799
7799
7799
<HgaI
TseI HaeIII
Fnu4HI CviJI
BthCI HaeI
>BbvI RsaI
MwoI HinP1I Csp6I
<CjePI HhaI >HgaI HpyCH4III Tsp509I
AGCAGGAAGCACTATGGGCGCAGCGTCAATGACGCTGACGGTACAGGCCAGACAATTATT 7860
TCGTCCTTCGTGATACCCGCGTCGCAGTTACTGCGACTGCCATGTCCGGTCTGTTAATAA
7806 7818 7831 7838 7854
7809 7818 7841
7820 7841
7820 7845
7820 7846
7820 7846
7823

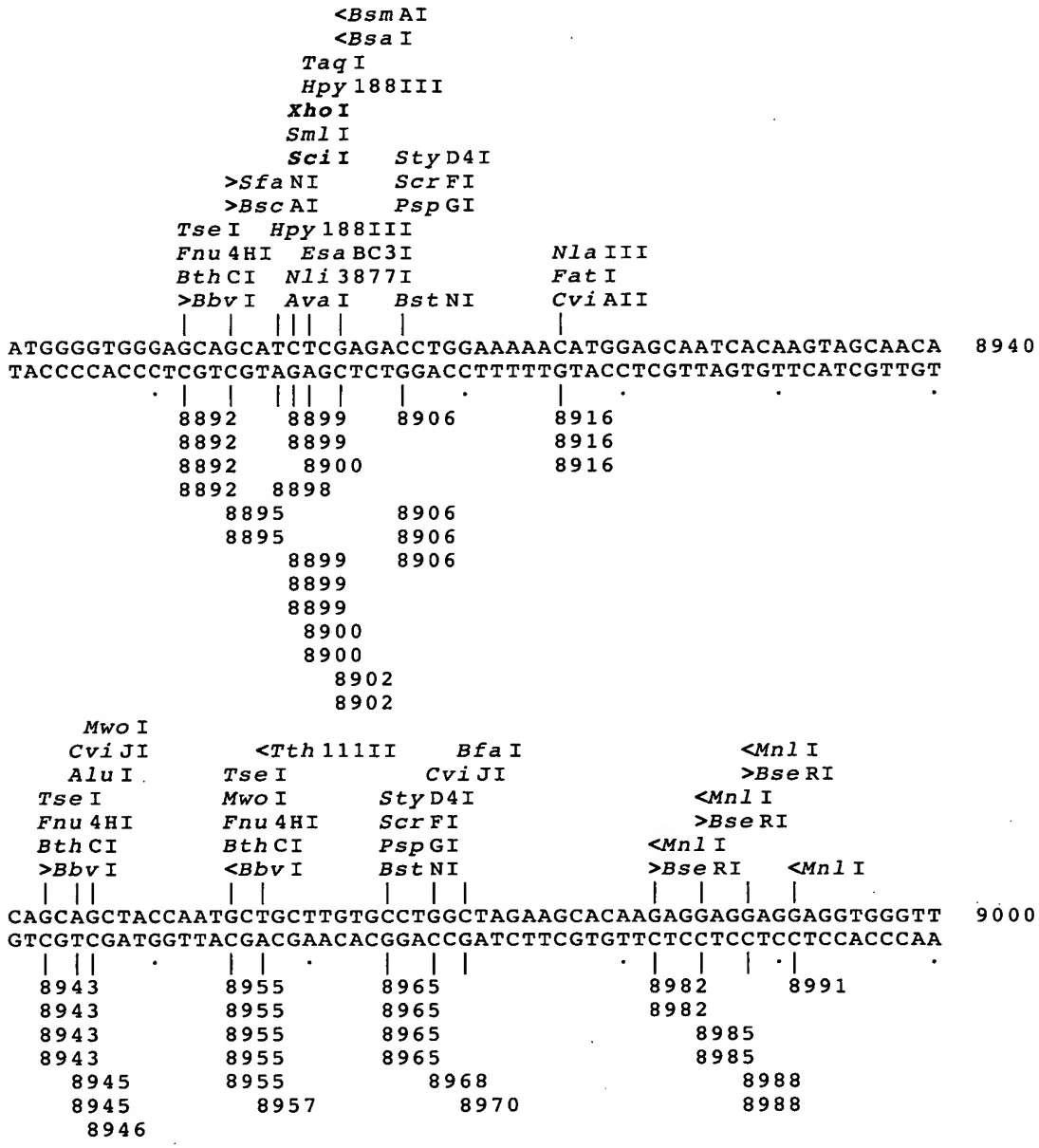

Tse I
Fnu 4HI
Bth CI
>Bbv I
<Mnl I
Tse I
Fnu 4HI
Bth CI
>Bbv I
Dde I
Mwo I
<Bsp CNI
<Bse MII
<Bpu 10I
<Bbv CI
>Sfa NI
Mwo I
Bst API
Hin P1I
>Bsc AI
Hha I
Apa BI
Hpy CH4V
>Bsg I
Tsp 509I
Cvi JI
<Mnl I
Alw NI
GTCTGGTATAGTGCAGCAGCAGAACAATTTGCTGAGGGCTATTGAGGCGCAACAGCATCT 7920
CAGACCATATCACGTCGTCGTCTTGTTAAACGACTCCCGATAACTCCGCGTTGTCGTAGA
7871
7872
7873
7873
7873
7873
7876
7876
7876
7876
7886
7891
7891
7892
7892
7892 7898
7894
7897
7904
7907
7907
7913
7915
7915
7915
7915
Sty D4I
Scr FI
Psp GI
<Eco 57MI
<Bpm I
Cvi JI
Tse I
Fnu 4HI
Bth CI
>Bbv I
Bst NI
Sty D4I
Scr FI
Psp GI
Bst NI
Bsl I
>Sfa NI
Alu I
Hpy CH4III
>Tth 111II
Tfi I
Hpy CH4V
>Bsc AI
>Ace III
Hin fI
Cvi JI
GTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAATCCTGGCTGTGGAAAG 7980
CAACGTTGAGTGTCAGACCCCGTAGTTCGTCGAGGTCCGTTCTTAGGACCGACACCTTTC
7923
7931
7941
7941
7945
7948
7948
7948
7948 7954
7949
7950
7950
7952
7952
7954
7954
7954
7962
7962
7966
7966
7966
7966
7966
7969

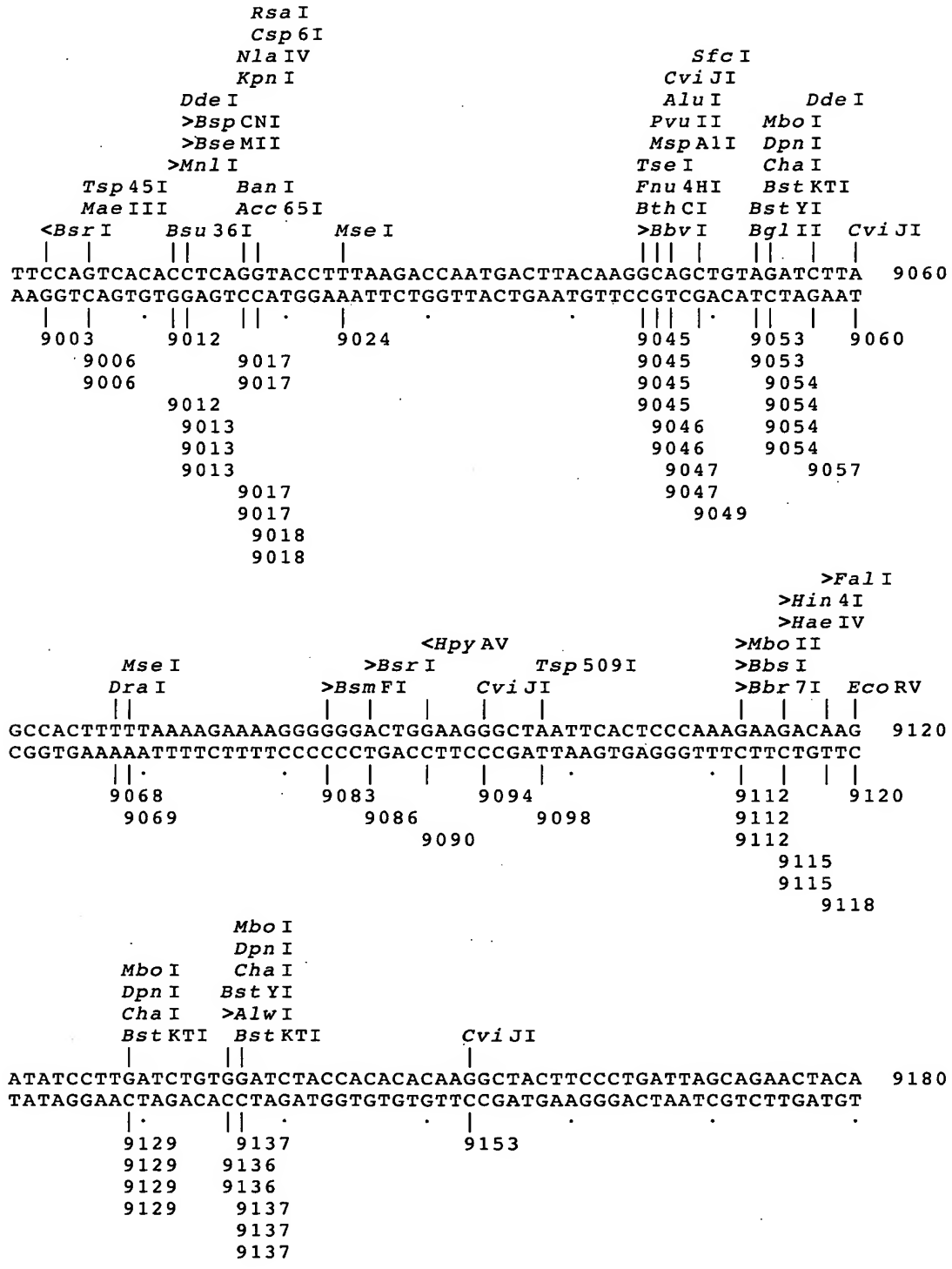

Sty D4I
Scr FI
Psp GI
Bst NI
<Cje I
Mbo I
Dpn I
Cha I
Bst KTI
Cvi JI
Alu I
Alw NI
Bsa JI
>Alw I
>Ace III
Hpy 188III
Hpy CH4V
ATACCTAAAGGATCAACAGCTCCTGGGGATTTGGGGTTGCTCTGGAAAACTCATTTGCAC 8040
TATGGATTTCCTAGTTGTCGAGGACCCCTAAACCCCAACGAGACCTTTTGAGTAAACGTG
7990
7997
8021
8036
7991
8002
7991
7997
7991
7998
7991
7998
7996
8002
8002
8002
8002
>Tsp RI
Msl I
>Bsm I
>Hin 4I
>Hae IV
<Cje I
<Bts I
Ale I
Sty I
Bsa JI
<Mme I
Bfa I
<Cje PI
Hpy 188III
Tfi I
Hin fI
CACTGCTGTGCCTTGGAATGCTAGTTGGAGTAATAAATCTCTGGAACAGATTTGGAATCA 8100
GTGACGACACGGAACCTTACGATCAACCTCATTATTTAGAGACCTTGTCTAAACCTTAGT
8041
8051
8061
8080
8095
8041
8051
8064
8084
8095
8041
8056
8089
8100
8041
8089
<Bcc I
>Sts I
>Fok I
>Bst F5I
Sty D4I
Scr FI
Psp GI
Bst NI
>Bsm FI
Mse I
Tsp 509I
Cvi JI
Alu I
Hin dIII
<Cst MI
CACGACCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAGCTTAATACACTCCTT 8160
GTGCTGGACCTACCTCACCCTGTCTCTTTAATTGTTAATGTGTTCGAATTATGTGAGGAA
8106
8118
8128
8143
8155
8106
8130
8144
8159
8106
8135
8144
8106
8147
8109
8109
8109
8110
Tfi I
Hin fI
>Mbo II
Tsp 509I
Mwo I
>Tsp DTI
Tsp 509I
Tsp 509I
AATTGAAGAATCGCAAAACCAGCAAGAAAAGAATGAACAAGAATTATTGGAATTAGATAA 8220
TTAACTTCTTAGCGTTTTGGTCGTTCTTTTCTTACTTGTTCTTAATAACCTTAATCTATT
8161
8173
8193
8202
8211
8165
8168
8168

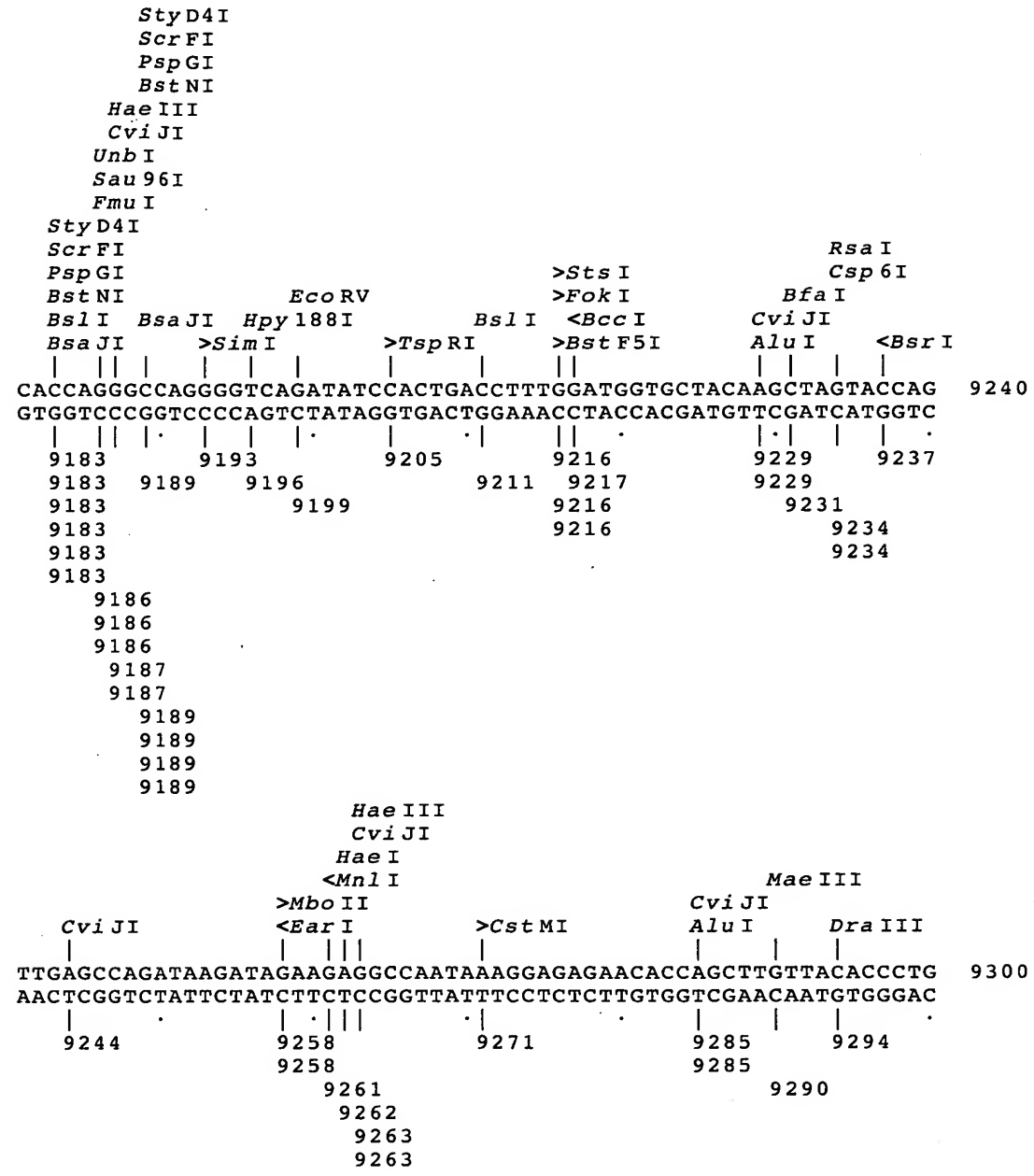

MseI
CviJI
<TspDTI
Tsp509I
Tsp509I
Tsp509I
ATGGGCAAGTTTGTGGAATTGGTTTAACATAACAAATTGGCTGTGGTATATAAAATTATT 8280
TACCCGTTCAAACACCTTAACCAAATTGTATTGTTTAACCGACACCATATATTTTAATAA
8237
8244
8255
8259
8274
8279
CviJI
<MnlI
<BsaXI
<BsaXI
MseI
RsaI
Csp6I
TatI
SfcI
CATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGCTGTACTTTCTATAGT 8340
GTATTACTATCATCCTCCGAACCATCCAAATTCTTATCAAAAACGACATGAAAGATATCA
8294
8294
8295
8297
8309
8326
8327
8327
8334
<MnlI
BsaJI
>Sth132I
NliI3877I
<SspD5I
<HphI
<SimI
Hpy188I
>MnlI
AvaI
GAATAGAGTTAGGCAGGGATATTCACCATTATCGTTTCAGACCCACCTCCCAACCCCGAG 8400
CTTATCTCAATCCGTCCCTATAAGTGGTAATAGCAAAGTCTGGGTGGAGGGTTGGGGCTC
8363
8363
8377
8380
8386
8395
8395
8395
8396
8398
>Sth132I
<SimI
VpaK11AI
UnbI
Sau96I
Psp03I
Nla IV
FmuI
AvaII
<HpyAV
SanDI
>Sth132I
PssI
UnbI
PpuMI
Sau96I
Nla IV
HaeIII
Eco0109I
FmuI
<HpyAV
>BsmFI
CviJI
>MboII
>MboII
<BsmAI
<BsmAI
BstYI
MboI
DpnI
ChaI
BstKTI
<AlwI
GGGACCCGACAGGCCCGAAGGAATAGAAGAAGAAGGTGGAGAGAGAGACAGAGACAGATC 8460
CCCTGGGCTGTCCGGGCTTCCTTATCTTCTTCTTCCACCTCTCTCTCTGTCTCTGTCTAG
8401
8401
8401
8401
8401
8401
8402
8402
8402
8402
8402
8402
8402
8403
8405
8412
8412
8412
8412
8412
8414
8417
8426
8429
8432
8445
8451
8456
8457
8457
8457
8457
8457

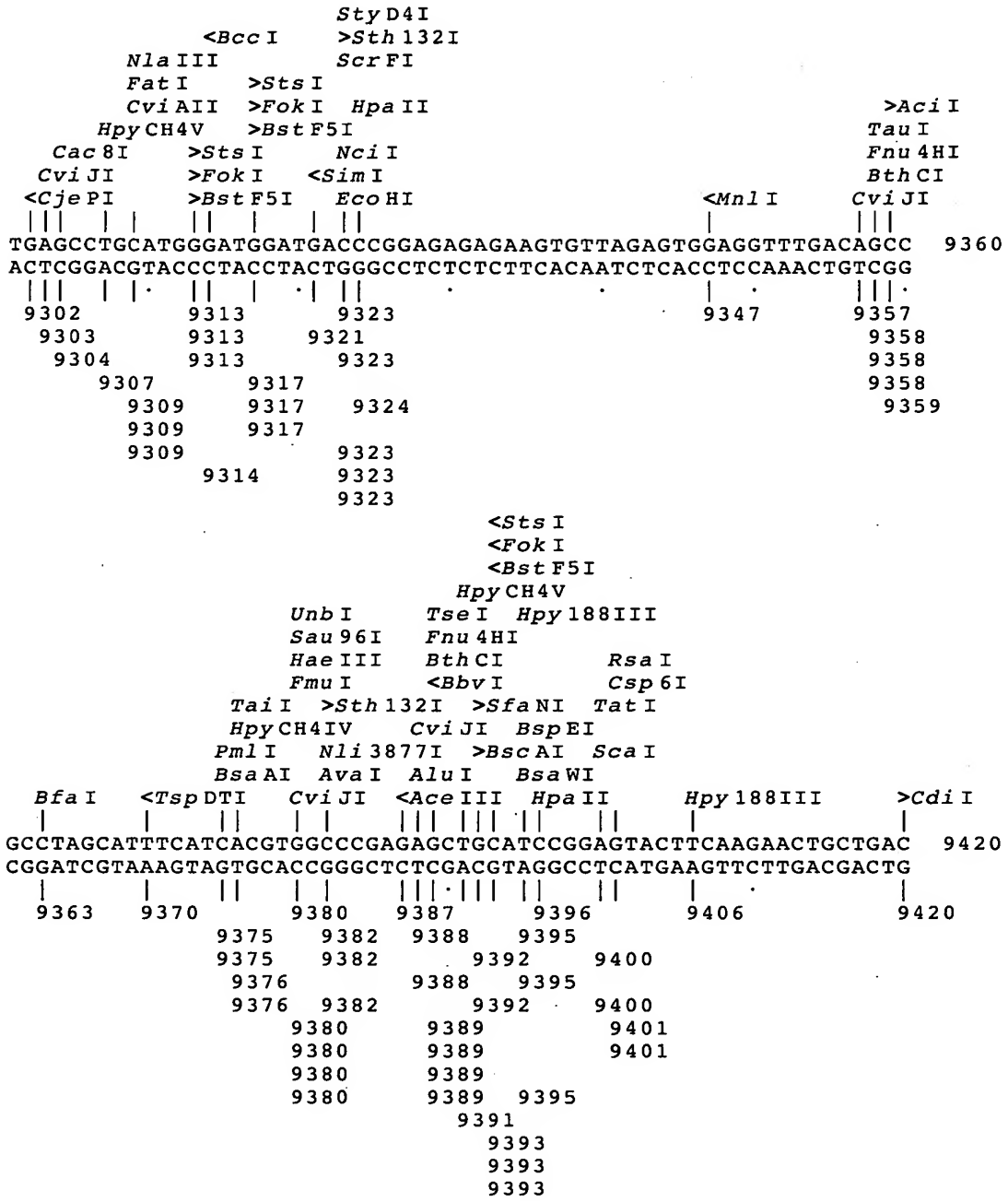

Sty I
Bsa JI
Mbo I
Dpn I
Cha I
Bst KTI
Nla IV
Bst YI
BamHI
>Alw I
Taq I
Hpy 8I
<Alw I
Esa BC3I
>Tsp GWI
Mbo I
Dpn I
Cha I
Nla IV
Bst KTI
Cvi JI
<Mbo II
>Ear I
>Mnl I
>Bsm FI
<Aci I
Mwo I
CATTCGATTAGTGAACGGATCCTTGGCACTTATCTGGGACGATCTGCGGAGCCTGTGCCT 8520
GTAAGCTAATCACTTGCCTAGGAACCGTGAATAGACCCTGCTAGACGCCTCGGACACGGA
8464 8475 8496 8506 8517
8464 8471 8478 8501 8510 8518
8477 8501 8508 8519
8477 8501 8520
8477 8501
8477
8478
8478
8478
8478
8481
8481
Mwo I
Cvi JI
Alu I
<Bsm AI
>Bpl I
Sml I
>Bpu EI
<Eco 57MI
<Acu I
>Aci I
>Bpl I
Hpy 188III
Mae III
<Mnl I
CTTCAGCTACCACCGCTTGAGAGACTTACTCTTGATTGTAACGAGGATTGTGGAACTTCT 8580
GAAGTCGATGGTGGCGAACTCTCTGAATGAGAACTAACATTGCTCCTAACACCTTGAAGA
8521 8533 8541 8550 8563
8521 8536 8558
8525 8536
8525 8541
8526 8541
>Ple I
>Mly I
>Hga I
>Bsm FI
>Mnl I
Cvi JI
Ssp I
Tfi I
Hin fI
Hpy CH4III
Sfc I
Hpy 188III
Hin fI
GGGACGCAGGGGGTGGGAAGCCCTCAAATATTGGTGGAATCTCCTACAGTATTGGAGTCA 8640
CCCTGCGTCCCCCACCCTTCGGGAGTTTATAACCACCTTAGAGGATGTCATAACCTCAGT
8581 8599 8607 8617 8624 8635
8583 8602 8617 8626 8638
8635
8635

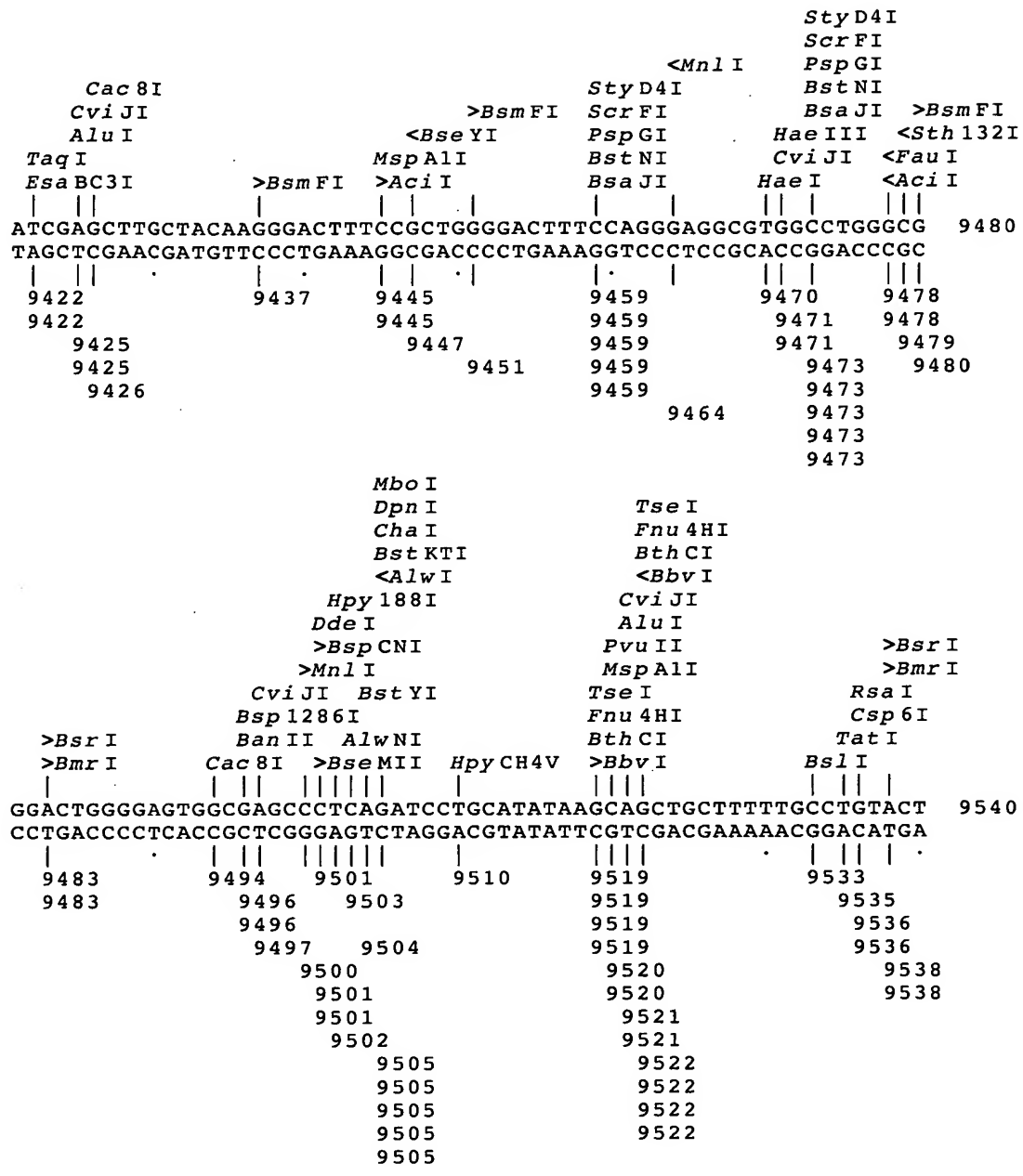

Dde I
<Bsp CNI
<Bse MII
<Bpu 10I
<Bbv CI
Cac 8I
Cvi JI
Alu I
Cvi JI
Alu I
Alw NI
>Bsm FI
<Mnl I
GGAACTAAAGAATAGTGCTGTTAGCTTGCTCAATGCCACAGCCATAGCAGTAGCTGAGGG 8700
CCTTGATTTCTTATCACGACAATCGAACGAGTTACGGTGTCGGTATCGTCATCGACTCCC
8663
8680
8688
8696
8663
8692
8699
8664
8692
8693
8693
8694
8694
8694
>Cst MI
Rsa I
Mwo I
Csp 6I
Cvi JI
Cvi JI
>Mbo II
Tat I
Alu I
Alu I
Bfa I
GACAGATAGGGTTATAGAAGTAGTACAAGGAGCTTGTAGAGCTATTCGCCACATACCTAG 8760
CTGTCTATCCCAATATCTTCATCATGTTCCTCGAACATCTCGATAAGCGGTGTATGGATC
8722
8731
8740
8757
8723
8731
8740
8760
8723
8732
8727
Cvi JI
<Cje I
<Cje PI
>Bsp 24I
<Bcc I
AAGAATAAGACAGGGCTTGGAAAGGATTTTGCTATAAGATGGGTGGCAAGTGGTCAAAAA 8820
TTCTTATTCTGTCCCGAACCTTTCCTAAAACGATATTCTACCCACCGTTCACCAGTTTTT
8769
8798
8769
8770
8774
Cvi JI
Dde I
Mwo I
Tse I
Blp I
Fnu 4HI
Cvi JI
Cvi JI
Bth CI
<Bcc I
Alu I
Cac 8I
>Sts I
<Ace III
>Bbv I
>Fok I
<Bsm BI
<Bsp CNI
>Bst F5I
Hpy CH4III
<Bsm AI
<Bse MII
<Bcc I
GTAGTGTGATTGGATGGCTTACTGTAAGGGAAAGAATGAGACGAGCTGAGCCAGCAGCAG 8880
CATCACACTAACCTACCGAATGACATTCCCTTTCTTACTCTGCTCGACTCGGTCGTCGTC
8832
8841
8858
8866
8880
8832
8858
8866
8832
8863
8874
8833
8864
8870
8836
8864
8874
8865
8874
8865
8874
8866
8869

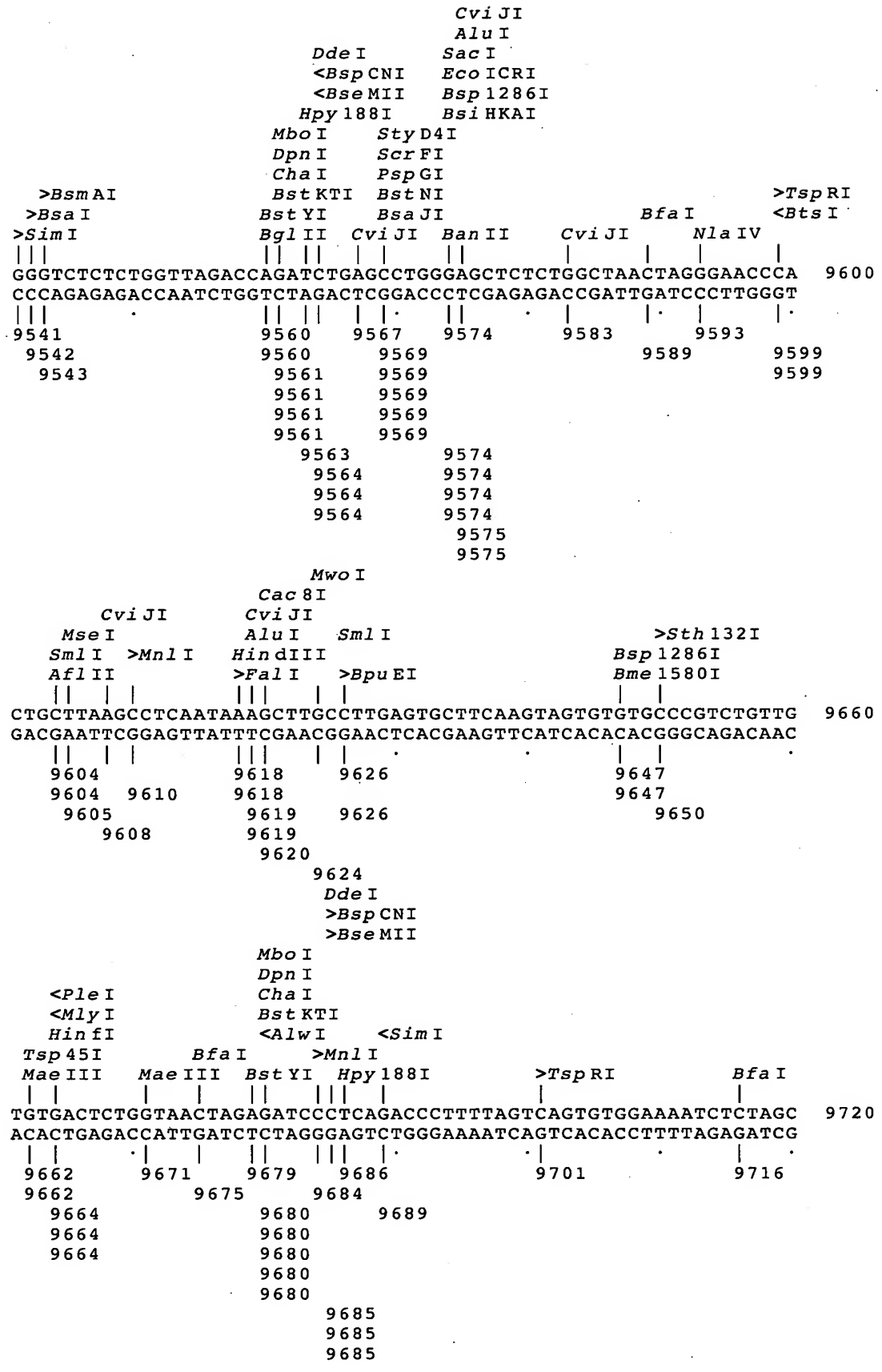

<Bsm AI
<Bsa I
Taq I
Hpy 188III
Xho I
Sml I
Sci I Sty D4I
>Sfa NI Scr FI
>Bsc AI Psp GI
Tse I Hpy 188III
Fnu 4HI Esa BC3I Nla III
Bth CI Nli 3877I Fat I
>Bbv I Ava I Bst NI Cvi AII
ATGGGGTGGGAGCAGCATCTCGAGACCTGGAAAAACATGGAGCAATCACAAGTAGCAACA 8940
TACCCCACCCTCGTCGTAGAGCTCTGGACCTTTTTGTACCTCGTTAGTGTTCATCGTTGT
8892 8899 8906 8916
8892 8899 8916
8892 8900 8916
8892 8898
8895 8906
8895 8906
8899 8906
8899
8899
8900
8900
8902
8902
Mwo I
Cvi JI <Tth 111II Bfa I <Mnl I
Alu I Tse I Cvi JI >Bse RI
Tse I Mwo I Sty D4I <Mnl I
Fnu 4HI Fnu 4HI Scr FI >Bse RI
Bth CI Bth CI Psp GI <Mnl I
>Bbv I <Bbv I Bst NI >Bse RI <Mnl I
CAGCAGCTACCAATGCTGCTTGTGCCTGGCTAGAAGCACAAGAGGAGGAGGAGGTGGGTT 9000
GTCGTCGATGGTTACGACGAACACGGACCGATCTTCGTGTTCTCCTCCTCCTCCACCCAA
8943 8955 8965 8982 8991
8943 8955 8965 8982
8943 8955 8965 8985
8943 8955 8965 8985
8945 8955 8968 8988
8945 8957 8970 8988
8946


Rsa I
Csp 6I
Nla IV
Kpn I
Dde I
>Bsp CNI
>Bse MII
>Mnl I
Tsp 45I
Mae III
Ban I
Acc 65I
<Bsr I
Bsu 36I
Mse I
Sfc I
Cvi JI
Alu I
Dde I
Pvu II
Mbo I
Msp A1I
Dpn I
Tse I
Cha I
Fnu 4HI
Bst KTI
Bth CI
Bst YI
>Bbv I
Bgl II
Cvi JI
TTCCAGTCACACCTCAGGTACCTTTAAGACCAATGACTTACAAGGCAGCTGTAGATCTTA 9060
AAGGTCAGTGTGGAGTCCATGGAAATTCTGGTTACTGAATGTTCCGTCGACATCTAGAAT
9003 9012 9024 9045 9053 9060
9006 9017 9045 9053
9006 9017 9045 9054
9012 9045 9054
9013 9046 9054
9013 9046 9054
9013 9047 9057
9017 9047
9017 9049
9018
9018
>Fal I
>Hin 4I
>Hae IV
<Hpy AV
>Mbo II
Mse I
>Bsr I
Tsp 509I
>Bbs I
Dra I
>Bsm FI
Cvi JI
>Bbr 7I
Eco RV
GCCACTTTTTAAAAGAAAAGGGGGGACTGGAAGGGCTAATTCACTCCCAAAGAAGACAAG 9120
CGGTGAAAAATTTTCTTTTCCCCCCTGACCTTCCCGATTAAGTGAGGGTTTCTTCTGTTC
9068 9083 9094 9112 9120
9069 9086 9098 9112
9090 9112
9115
9115
9118
Mbo I
Dpn I
Cha I
Bst YI
>Alw I
Bst KTI
Bst KTI
Cvi JI
ATATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCCTGATTAGCAGAACTACA 9180
TATAGGAACTAGACACCTAGATGGTGTGTGTTCCGATGAAGGGACTAATCGTCTTGATGT
9129 9137 9153
9129 9136
9129 9136
9129 9137
9137
9137


StyD4I
ScrFI
PspGI
BstNI
HaeIII
CviJI
UnbI
Sau96I
FmuI
StyD4I
ScrFI
PspGI
BstNI
BslI
BsaJI
BsaJI
EcoRV
Hpy188I
>SimI
>TspRI
BslI
>StsI
>FokI
<BccI
>BstF5I
RsaI
Csp6I
BfaI
CviJI
AluI
<BsrI
CACCAGGGCCAGGGGTCAGATATCCACTGACCTTTGGATGGTGCTACAAGCTAGTACCAG 9240
GTGGTCCCGGTCCCCAGTCTATAGGTGACTGGAAACCTACCACGATGTTCGATCATGGTC
9183 9193 9205 9216 9229 9237
9183 9189 9196 9211 9217 9229
9183 9199 9216 9231
9183 9216 9234
9183 9234
9183
9186
9186
9186
9187
9187
9189
9189
9189
9189
HaeIII
CviJI
HaeI
<MnlI
MaeIII
>MboII
CviJI
CviJI
<EarI
>CstMI
AluI
DraIII
TTGAGCCAGATAAGATAGAAGAGGCCAATAAAGGAGAGAACACCAGCTTGTTACACCCTG 9300
AACTCGGTCTATTCTATCTTCTCCGGTTATTTCCTCTCTTGTGGTCGAACAATGTGGGAC
9244 9258 9271 9285 9294
9258 9285
9261 9290
9262
9263
9263


Sty D4I
<Bcc I
>Sth 132I
Nla III
Scr FI
Fat I
>Sts I
Cvi AII
>Fok I
Hpa II
>Aci I
Hpy CH4V
>Bst F5I
Tau I
Cac 8I
>Sts I
Nci I
Fnu 4HI
Cvi JI
>Fok I
<Sim I
Bth CI
<Cje PI
>Bst F5I
Eco HI
<Mnl I
Cvi JI
TGAGCCTGCATGGGATGGATGACCCGGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCC 9360
ACTCGGACGTACCCTACCTACTGGGCCTCTCTCTTCACAATCTCACCTCCAAACTGTCGG
9302 9313 9323 9347 9357
9303 9313 9321 9358
9304 9313 9323 9358
9307 9317 9358
9309 9317 9324 9359
9309 9317
9309 9323
9314 9323
9323
<Sts I
<Fok I
<Bst F5I
Hpy CH4V
Unb I
Tse I
Hpy 188III
Sau 96I
Fnu 4HI
Hae III
Bth CI
Rsa I
Fmu I
<Bbv I
Csp 6I
Tai I
>Sth 132I
>Sfa NI
Tat I
Hpy CH4IV
Cvi JI
Bsp EI
Pml I
Nli 3877I
>Bsc AI
Sca I
Bsa AI
Ava I
Alu I
Bsa WI
Bfa I
<Tsp DTI
Cvi JI
<Ace III
Hpa II
Hpy 188III
>Cdi I
GCCTAGCATTTCATCACGTGGCCCGAGAGCTGCATCCGGAGTACTTCAAGAACTGCTGAC 9420
CGGATCGTAAAGTAGTGCACCGGGCTCTCGACGTAGGCCTCATGAAGTTCTTGACGACTG
9363 9370 9380 9387 9396 9406 9420
9375 9382 9388 9395
9375 9382 9392 9400
9376 9388 9395
9376 9382 9392 9400
9380 9389 9401
9380 9389 9401
9380 9389
9380 9389 9395
9391
9393
9393
9393


Sty D4I
Scr FI
<Mnl I
Psp GI
Bst NI
Bsa JI
>Bsm FI
Hae III
<Sth 132I
Cvi JI
<Fau I
Hae I
<Aci I
Cac 8I
Cvi JI
Alu I
Taq I
Esa BC3I
>Bsm FI
<Bse YI
Msp A1I
>Aci I
Sty D4I
Scr FI
Psp GI
Bst NI
Bsa JI
ATCGAGCTTGCTACAAGGGACTTTCCGCTGGGGACTTTCCAGGGAGGCGTGGCCTGGGCG 9480
TAGCTCGAACGATGTTCCCTGAAAGGCGACCCCTGAAAGGTCCCTCCGCACCGGACCCGC
9422
9422
9425
9425
9426
9437
9445
9445
9447
9451
9459
9459
9459
9459
9464
9470
9471
9471
9473
9473
9473
9473
9478
9478
9479
9480
Mbo I
Dpn I
Cha I
Bst KTI
<Alw I
Hpy 188I
Dde I
>Bsp CNI
>Mnl I
Cvi JI
Bst YI
Bsp 1286I
Ban II
Alw NI
Tse I
Fnu 4HI
Bth CI
<Bbv I
Cvi JI
Alu I
Pvu II
Msp A1I
Tse I
Fnu 4HI
Bth CI
>Bbv I
>Bsr I
>Bmr I
Rsa I
Csp 6I
Tat I
Bsl I
>Bsr I
>Bmr I
Cac 8I
>Bse MII
Hpy CH4V
GGACTGGGGAGTGGCGAGCCCTCAGATCCTGCATATAAGCAGCTGCTTTTTGCCTGTACT 9540
CCTGACCCCTCACCGCTCGGGAGTCTAGGACGTATATTCGTCGACGAAAAACGGACATGA
9483
9483
9494
9496
9496
9497
9500
9501
9501
9502
9505
9505
9505
9505
9505
9501
9503
9504
9510
9519
9519
9519
9519
9520
9520
9521
9521
9522
9522
9522
9522
9533
9535
9536
9536
9538
9538


CviJI
AluI
DdeI
SacI
<BspCNI
EcoICRI
<BseMII
Bsp1286I
Hpy188I
BsiHKAI
MboI
StyD4I
DpnI
ScrFI
ChaI
PspGI
>BsmAI
BstKTI
BstNI
>TspRI
>BsaI
BstYI
BsaJI
BfaI
<BtsI
>SimI
BglII
CviJI
BanII
CviJI
NlaIV
GGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCA 9600
CCCAGAGAGACCAATCTGGTCTAGACTCGGACCCTCGAGAGACCGATTGATCCCTTGGGT
9541 9560 9567 9574 9583 9593
9542 9560 9569 9589 9599
9543 9561 9569 9599
9561 9569
9561 9569
9561 9569
9563 9574
9564 9574
9564 9574
9564 9574
9575
9575
MwoI
CacI8I
CviJI
CviJI
MseI
AluI
SmlI
>Sth132I
SmlI
>MnlI
HindIII
Bsp1286I
AflII
>FalI
>BpuEI
Bme1580I
CTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTG 9660
GACGAATTCGGAGTTATTTCGAACGGAACTCACGAAGTTCATCACACACGGGCAGACAAC
9604 9618 9626 9647
9604 9610 9618 9647
9605 9619 9626 9650
9608 9619
9620
9624
DdeI
>BspCNI
>BseMII
MboI
DpnI
<PleI
ChaI
<MlyI
BstKTI
HinfI
<AlwI
<SimI
Tsp45I
BfaI
>MnlI
MaeIII
MaeIII
BstYI
Hpy188I
>TspRI
BfaI
TGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGC 9720
ACACTGAGACCATTGATCTCTAGGGAGTCTGGGAAAATCAGTCACACCTTTTAGAGATCG
9662 9671 9679 9686 9701 9716
9662 9675 9684
9664 9680 9689
9664 9680
9664 9680
9680
9680
9685
9685
9685

A 9721
T

///

JOURNAL OF VIROLOGY, May 1991, p. 2415–2421
0022-538X/91/052415-07$02.00/0
Copyright © 1991, American Society for Microbiology
Vol. 65, No. 5

A Single-Stranded Gap in Human Immunodeficiency Virus Unintegrated Linear DNA Defined by a Central Copy of the Polypurine Tract

PIERRE CHARNEAU AND FRANÇOIS CLAVEL*

Unité d'Oncologie Virale, Département SIDA et Rétrovirus, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris Cedex 15, France

Received 9 November 1990/Accepted 7 February 1991

The structure of unintegrated human immunodeficiency virus type 1 (HIV-1) DNA from acutely infected human lymphoid cells was analyzed by nuclease S1 cleavage. We observed a unique, discrete single-stranded gap in unintegrated linear DNA molecules, located near the center of the genome. Oligonucleotide primer extension experiments determined that the downstream limit of this gap coincides with the last nucleotide of a central copy of the polypurine tract found in all sequenced lentivirus genomes. Other retroviruses have only one copy of the polypurine tract at the 5′ boundary of the 3′ long terminal repeat, which has been shown to determine initiation of retroviral DNA plus-strand synthesis. We conclude from our observations that the central repeat of the polypurine tract can create an additional site for plus-strand synthesis initiation in lentiviruses. The central single-stranded gap was not found in circular DNA molecules, the vast majority of them carrying only one long terminal repeat. This finding suggests that the generation of such circular molecules is associated with early DNA ligation events.

Retroviruses replicate through reverse transcription of their RNA genomes into a double-stranded DNA molecule. Retroviral genes are expressed from an integrated copy of this double-stranded DNA genome, the provirus. Both strands of the retroviral DNA genome are synthesized by the virus-encoded reverse transcriptase, which has both RNA- and DNA-dependent DNA polymerase activities (10, 38, 41). The template for minus-strand synthesis is viral genomic RNA, and the template for plus-strand synthesis is the newly reverse-transcribed minus strand, following removal of RNA from the RNA-DNA hybrid by the RNase H activity associated with reverse transcriptase (8, 11). Synthesis of the minus strand and subsequently of the plus strand is initiated near the 5′ end of the respective template. The corresponding short segments (minus- and plus-strand strong-stop DNAs) are further transferred to the other end of the template, resulting in the formation of the long terminal repeats (LTRs), present at each end of the provirus (10, 30, 37). The primer for minus-strand synthesis is the 3′ end of a tRNA molecule packaged in the viral particle together with genomic viral RNA and hybridized to the primer binding site located at the 3′ boundary of the U5 region (35, 36, 39). The initiation site of the plus strand has been deduced from analysis of reverse transcription reactions products obtained either in vitro or from detergent-disrupted virions and from the sequences of proviral molecular clones. This site is located immediately 3′ of a polypurine tract (PPT) representing the 5′ boundary of the U3 region (21, 22) (Fig. 1). It has been proposed that the PPT is used to define an RNA primer by specific cleavage of the RNA template at this site by the reverse transcriptase-associated RNase H (20, 23, 27, 32). Indeed, in murine retroviruses, in vitro reverse transcription reactions reveal RNA primers that remain associated with the elongating plus strand and are heterogeneous in length (7, 26). Unlike most retroviruses, human immunodeficiency viruses (HIV) and other lentiviruses have two copies of the PPT, one at the border of the 3′ LTR and the other located near the middle of the genome, within the *pol* coding region (5, 12, 33, 43). Previous experiments on visna virus, the prototypic lentivirus, have shown that unintegrated DNA molecules display a single-stranded gap located approximately in the same area (3, 14). Other viral DNA genomes carry single-stranded gaps or nicks. This is the case for hepatitis B virus and cauliflower mosaic virus, in which synthesis of viral DNA genome involves a reverse transcription step.

We have tested here the hypothesis that the HIV central PPT represents an additional initiation site for the synthesis of the plus strand of HIV DNA. We show that HIV linear unintegrated DNA molecules carry a discrete single-stranded gap whose downstream limit is the last nucleotide of the central PPT, indicating that this structure is likely used to initiate plus-strand synthesis at the center of the genome. We also show that the single-stranded gap is absent from circular molecules. Since it is known that such circular molecules are formed only after transport of reverse transcription products into the nucleus, we propose that nuclear ligation and DNA repair events result in both the closing of circular DNA molecules and filling of the gap.

MATERIALS AND METHODS

Cells and viruses. MT4 cells were a gift from M. David Hogan, Laboratory of Molecular Microbiology, National Institute of Allergy and Infectious Diseases, Bethesda, Md. These cells, which are transformed by human T-cell leukemia virus type I, were shown to allow acute cytopathic HIV-1 infection (13). CEM clone 13 cells (28) were derived from the human lymphoid cell line CEM (ATCC CCL119) and express high levels of CD4 antigen. Cells were maintained in RPMI 1640 medium (GIBCO Laboratories) supplemented with 10% fetal calf serum.

The viral isolate used in our experiments was HIV-1_{bru} (2),

* Corresponding author.

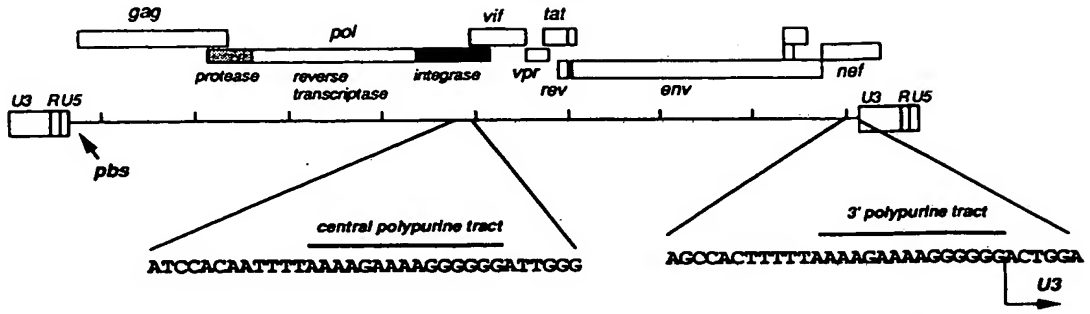


FIG. 1. Positions of the two PPTs on the HIV-1 genome. Shaded areas in the *pol* open reading frame indicate the regions coding for different functions. pbs, Primer binding site, representing the initiation site for minus-strand synthesis.

recovered following transfection of COS cells with infectious proviral molecular clone pBRU-2 (24a). Cells were infected at a multiplicity of 1:10 (1 50% tissue culture infective dose per 10 cells) with virus from a frozen (−80°C) stock produced on MT4 cells and titrating 8×10^5 tissue culture infective doses per ml on MT4 cells. Following infection, cultures were monitored for cytopathic effect, which in the described conditions of infection appeared approximately 3 days after infection. At that time, cells were harvested for DNA isolation.

Analysis of viral DNA. Low-molecular-weight DNA was extracted from infected cells by Hirt extraction (15). Nuclease S1 (Appligene, Strasbourg, France) was used at 1.5 U/μg of DNA after addition of 1:10 volume of 10× S1 buffer (300 mM sodium acetate [pH 4.6], 500 mM NaCl, 10 mM $ZnCl_2$) and incubated at 37°C. For double digestions with a restriction enzyme and nuclease S1, DNA (10 μg) was first digested with the restriction enzyme; then 10× nuclease S1 buffer and nuclease S1 (15 U) were added to the reaction mixture, and the sample was further incubated at 37°C. DNA was then subjected to electrophoresis on 1% agarose gels that did not contain ethidium bromide and analyzed by Southern blotting (34).

The nucleotide position numbers used here to describe DNA fragments start at the first nucleotide of a linear HIV DNA genome (5′ end of the U3 region in the 5′ LTR).

Two probes were used in hybridization experiments. The 5′ probe was a *PstI* fragment spanning the *gag* region from positions 1415 to 2839, and the 3′ probe was a *KpnI* fragment spanning the *env* region from positions 6343 to 9005. Both probes were obtained from HIV-1 molecular clone pNL4-3 (1) and labeled by the random hexamer method (6).

Primer extension. A modification of the primer extension technique was used, based on *Taq* DNA polymerase, which allows multiple cycles of denaturation, annealing, and polymerization. The reaction mix contained 10 μg of Hirt DNA, 0.2 mM each deoxynucleotide, 1.25 U of *Taq* DNA polymerase (Perkin-Elmer Cetus N801-0045), 10 pmol of 5′-end-labeled primer, and 0.5 μl of PMPE reagent (Stratagene) in 10 mM Tris-HCl (pH 8.3)–50 mM KCl–2 mM $MgCl_2$–0.01% gelatin, for a total reaction volume of 50 μl. The first cycle included 30 s for denaturation at 92°C, 1 min for annealing at 50°C, and 1 min for polymerization at 72°C. The following 59 cycles included 10 s at 92°C, 1 min at 50°C, and 1 min at 72°C. As only one primer was included in the reaction, this multicycle primer extension method is distinct from a polymerase chain reaction and resulted in linear, not exponential, amplification of the reaction product.

Oligonucleotide primers were 5′ end labeled with T4 polynucleotide kinase in the presence of [γ-^{32}P]ATP to a specific activity of 5×10^5 to 10^6 cpm/10 pmol. Primer pol-1 (5′-ACA ATC ATC ACC TGC CAT CTG) anneals to the plus strand at position 5085. Primer pol-2 (5′-TCC AAA GTG GAT CTC TGC TGT) anneals to the plus strand at position 4951.

The same oligonucleotide primers were used to generate a sequence ladder from a single-stranded M13 template carrying the $HIV\text{-}1_{bru}$ plus strand from positions 4688 to 5129. Sequencing was done by the method of Sanger et al. (29). Primer extension and sequence reactions were analyzed in parallel on a 6% polyacrylamide–8 M urea sequencing gel.

RESULTS

A single-stranded gap in HIV-1 unintegrated DNA. We hypothesized that the existence of an additional plus-strand initiation site in the HIV-1 genome could be revealed by the presence of a single-stranded region in HIV-1 unintegrated DNA molecules. To examine this possibility, low-molecular-weight DNA was selectively extracted (15) from acutely HIV-1-infected MT4 cells (13), treated with nuclease S1, and analyzed by Southern blotting with a probe spanning most of the 3′ half of the HIV-1 genome. Figure 2 shows the kinetics of nuclease S1 digestion of HIV-1 unintegrated DNA. In undigested DNA, three bands were observed. The position of the middle band, around 9.5 kb, is consistent with linear full-length molecules, and the position of the lower band, around 6 kb, is consistent with closed (supercoiled) circles. The upper band, which has an apparent size of approximately 15 kb, is likely to be open (relaxed) circular molecules. After 5 min of digestion with nuclease S1 (15 U/10 μg of DNA), major changes occurred. The band corresponding to linear molecules turned into a doublet; the upper band had the size of intact linear molecules (9.5 kb), and the lower one, at approximately 9 kb, was consistent with the size of linearized circles with one LTR. Simultaneously, the closed-circles 6-kb band completely disappeared, while the intensity of the open-circles band increased slightly. More strikingly, a new 5-kb band appeared, corresponding to the 3′ half of linear molecules cut into two fragments by nuclease S1. The other fragment, of similar size, could be detected in a distinct hybridization experiment with a 5′ probe (data not shown). This finding indicates the presence of a unique nuclease S1-sensitive site in HIV unintegrated DNA molecules located approximately at the center of the genome.

The complete disappearance of the closed circles, even

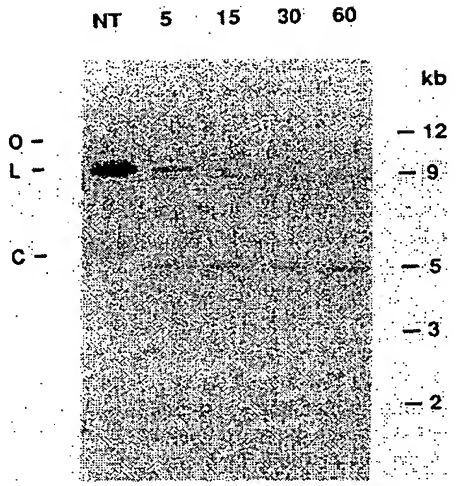


FIG. 2. Kinetics of nuclease S1 action on HIV-1 unintegrated DNA. Shown is a Southern blot of low-molecular-weight DNA from HIV-1-infected MT4 cells, hybridized to a probe representing most of the 3′ half of the HIV-1 genome (see Fig. 3B). Lanes: NT, no nuclease S1 treatment; 5, 15, 30, and 60, treatment with S1 nuclease (1.5 U/μg of DNA) for 5, 15, 30, and 60 min, respectively. O, Open circles, L, linear molecules; C, closed circles.

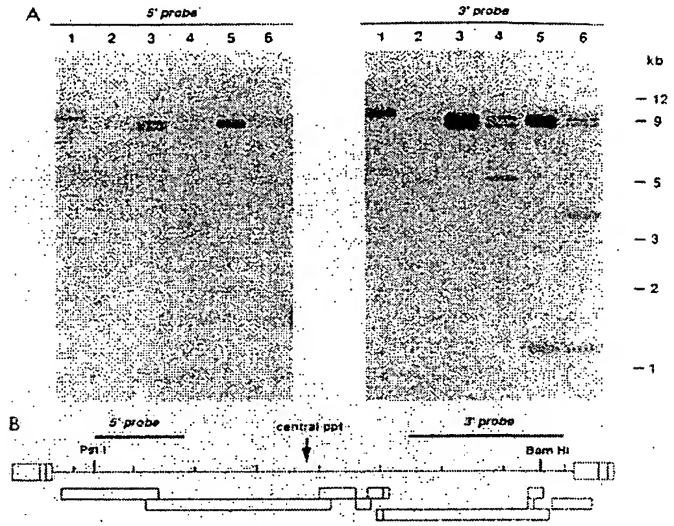


FIG. 3. Position of the single-stranded gap on the HIV-1 genome. (A) Two autoradiograms from the same Southern blot, hybridized to two different probes as indicated at the top. Low-molecular-weight DNA from HIV-1-infected MT4 cells was analyzed by nuclease S1 treatment either with no previous restriction enzyme digestion (lanes 1 and 2), after digestion with *Pst*I (lanes 3 and 4), or after digestion with *Bam*HI (lanes 5 and 6). Lanes 1, 3, and 5, No nuclease S1 treatment; lanes 2, 4, and 6, treatment with nuclease S1 (1.5 U/μg of DNA). (B) Positions of the corresponding probes on the HIV-1 genome, relative to the central PPT and to the unique *Pst*I and *Bam*HI restriction sites.

after 5 min of nuclease S1 treatment, can be best explained by a previously described high sensitivity of supercoiled circular DNA molecules to nuclease S1 (19). Therefore, we assumed that at this stage of incomplete nuclease S1 digestion, the closed circles were cleaved at random sites, into open circles if only one strand was cut (this explains the increase in open circles) and into linear molecules when S1 nuclease further cleaved at the resulting nicks.

With further digestion, the open-circles band gradually disappeared, together with the full-length linear molecules, while the band corresponding to linearized one-LTR circles remained unaffected. The gradual disappearance of open circles was due to their cleavage at random nicks that relate to their relaxed physical state: S1 cleavage at those sites linearized them. Meanwhile, linear molecules were further cleaved at the central S1-sensitive site, generating more of the probe-reactive 5-kb fragment: after 1 h of treatment, very little of the full-length linear molecules remained.

The central single-stranded gap is discrete and unique. We further attempted to localize and define the structure of the HIV-1 central nuclease S1-sensitive site. Figure 3A shows a Southern blot analysis of nuclease S1 cleavage products of HIV-1 unintegrated DNA, obtained from acutely HIV-1-infected MT4 cells, combined with digestion by two different restriction enzymes. In these experiments, complete restriction enzyme digestions were performed, followed by partial nuclease S1 digestion (20 min with 15 U/10 μg of DNA). The digestion products were examined by hybridization to two different HIV-1 probes: a 5′ probe, in the *gag* region, spanning from nucleotide positions 1415 (from the start of U3) to 2839, and a 3′ probe, spanning most of the *env* and *nef* regions, from nucleotide positions 6343 to 9005.

In the HIV-1_{bru} genome, *Pst*I has a unique site in the *gag* coding region at position 1415. Digestion of HIV-1 unintegrated DNA with *Pst*I (Fig. 3A, lanes 3) cut the linear molecules in two fragments of 8.3 and 1.4 kb, the latter being undetected by either of the two probes used in this experiment. *Pst*I also linearizes circles into a 9-kb product. Nuclease S1 treatment of *Pst*I digestion products (lanes 4) yielded a 5-kb fragment reactive with the 3′ probe and a 3.3-kb fragment reactive with the 5′ probe. This result establishes that both boundaries of the central single-stranded gap are discrete.

After digestion with *Bam*HI (Fig. 3A, lanes 5), which has a unique site in the *env* region at position 8520, the linear DNA molecules were cleaved into two fragments of approximately 8.5 and 1.2 kb, both reactive with the 3′ probe. Treatment with nuclease S1 (lanes 6) released a 3.7-kb fragment detected by the 3′ probe and a 4.9-kb fragment reactive with the 5′ probe. The 1.2-kb fragment, which includes the 3′ PPT, remained unaffected by nuclease S1 treatment. When no restriction enzyme digestion was performed before nuclease S1 treatment, S1 cleavage (lanes 2) released a 5′ 4.9-kb fragment and a 3′ fragment of similar size. Overall, the results from these single- and double-digestion experiments were consistent and could locate the single-stranded gap around position 4900, within 0.2 kb of the central PPT repeat found in the HIV-1_{bru} sequence.

These experiments show that the central single-stranded gap is unique in the HIV-1 genome. Indeed, the sizes of the fragments released by both the single and double digestions excluded the presence of another gap.

The 3′ boundary of the gap is the central copy of the PPT. To locate the 3′ boundary of the single-stranded gap more precisely relative to the central PPT found in all lentiviral genomes, primer extension experiments were conducted, using low-molecular-weight DNA from acutely HIV-1-infected lymphoid cells as a template. We used oligonucleotide primers complementary to the viral plus strand, located

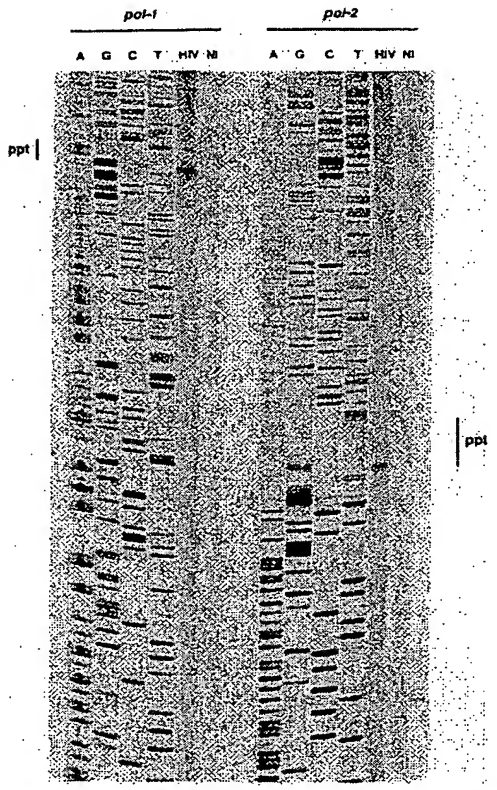


FIG. 4. Localization of the downstream limit of the single-stranded gap by oligonucleotide primer extension. Two primer extension reactions with two different oligonucleotide primers are presented. Primer pol-1 anneals to the HIV-1 plus strand at positions 5085 to 5105, and primer pol-2 anneals at positions 4951 to 4971. Both oligonucleotides were ^{32}P 5′ end labeled and allowed to prime an extension reaction, using low-molecular-weight DNA from HIV-1-infected cells (lanes HIV) or noninfected cells (lanes NI) as templates as described in Materials and Methods. Lanes A, G, C, and T are sequence reactions from a single-stranded M13 template carrying the HIV-1_{bru} plus strand from positions 4688 to 5129, using the indicated oligonucleotide as a primer. ppt, Central copy of the PPT.

downstream to the central copy of the PPT, expecting elongation of these primers to stop precisely at any interruption of the plus strand. To increase the sensitivity, the enzyme used was *Taq*I polymerase, which allowed us to heat denature extension reaction mixes and to carry several cycles of extension for each reaction. Two different antisense 5′-end-labeled oligonucleotide primers were used. Primer pol-2 anneals to the plus-strand 135 nucleotides downstream of the PPT, and primer pol-1 anneals 134 nucleotides further downstream. In parallel with extension reactions, DNA sequence reactions were conducted, using the same oligonucleotide primers on single-stranded M13–HIV-1_{bru} templates, enabling us to precisely locate the stop in primer extension on the HIV-1_{bru} genome (Fig. 4). With both primers we observed a single, discrete stop in primer extension that coincided with the last nucleotide of the central PPT. No equivalent stop was found with low-molecular-weight DNA from uninfected cells. In addition, no HIV-specific signal could be observed with a primer complementary to the minus strand, located upstream of the central PPT (data not shown). This result demonstrates that the 3′ boundary of the HIV-1 central single-stranded gap is defined by the central copy of the PPT. We can infer from this observation that the central PPT is used as an additional initiation site for the synthesis of the plus strand of HIV-1 DNA.

The single-stranded gap is found exclusively on linear DNA molecules. Electrophoretic analysis of undigested and of *Pst*I- or *Bam*HI-digested unintegrated HIV-1 DNA showed an approximately equal proportion of linear and circular molecules in acutely HIV-infected MT4 cells. Indeed, after *Pst*I or *Bam*HI digestion, the intensity of the shortened product resulting from digestion of linear molecules was approximately equal to that of the 9-kb linearized circles, found in both *Bam*HI and *Pst*I digestion reactions. Of interest is that no band with a size similar to that of the full-length linear molecules (corresponding to two-LTR linearized circles) could be observed in these reactions. This means that the vast majority of circular molecules in the acutely HIV-1-infected MT4 system have only one LTR.

Several lines of evidence indicate that circular molecules do not carry a single-stranded gap. First, the closed circles by definition cannot be gapped, although they are highly sensitive to nuclease S1 digestion (19). Second, the S1 digestion kinetics experiment (Fig. 2) shows that the 9-kb product corresponding to the linearized one-LTR circles was still present, even after 1 h of treatment with nuclease S1, when native linear molecules had almost completely disappeared. Third, after digestion with *Pst*I and *Bam*HI, which cut only once in the HIV-1_{bru} genome, nuclease S1 treatment did not generate products of a size compatible with fragments released from gapped circular molecules. For example, the *Bam*HI-nuclease S1 double digestion failed to release a 5.5-kb fragment, which would span from the *Bam*HI site to the gap in a circular one-LTR molecule and would be reactive with the 5′ probe. Similarly, the *Pst*I-nuclease S1 double digestion did not release any 3′-probe-reactive product of 5.7 kb, which would be the distance between the gap site and the *Pst*I site in a circular one-LTR molecule. Overall, the products observed in the double digests were of a size compatible only with DNA fragments released from native, linear two-LTR molecules.

Finally, we have observed that the proportion of circular versus linear molecules could vary with the source of the analyzed DNA. Figure 5 shows a comparison of nuclease S1 treatment of low-molecular-weight DNA from MT4 cells harvested 2 days after HIV-1 infection (lanes 1 to 4) and from CEM cells harvested 9 days after infection (lanes 5 to 8). In the CEM cells, the proportion of linear molecules was low, as shown by the lower proportion of 9.5-kb band than of open and closed circles on undigested lane 5. Another sign of the low proportion of linear molecules is visible on lane 7, where the intensity of the 1.2-kb fragment released by *Bam*HI is weaker than that of the 9-kb linearized circles. Coincidentally, the amount of subgenomic product released by nuclease S1 digestion of the same DNA samples was low. In the MT4 cells, in which more linear molecules were found, the quantity of subgenomic S1 digestion product was proportionally much higher than in the CEM cells.

DISCUSSION

We have observed that HIV-1 unintegrated linear DNA molecules, which accumulate in acutely infected CD4+ lymphoid cells, carry a short, unique, single-stranded region at the center of the genome. Both ends of this single-

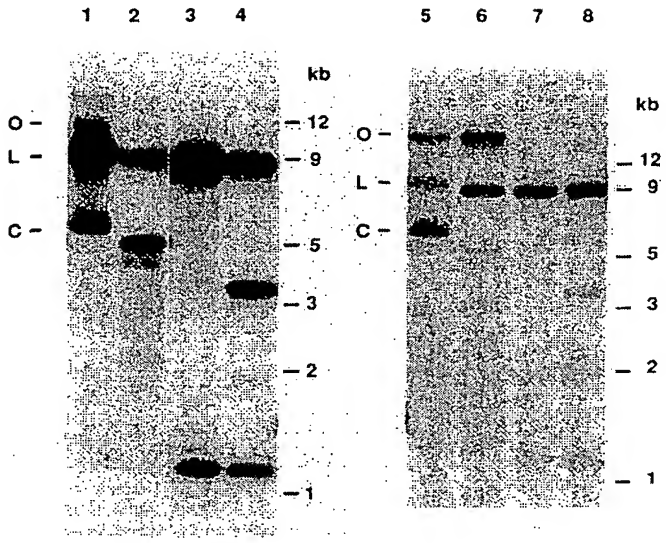


FIG. 5. Nuclease S1 sensitivity of HIV-1 unintegrated DNA from two different cell cultures. Lanes: 1 to 4, low-molecular-weight DNA from MT4 cells harvested 2 days after infection; 5 to 8, low-molecular-weight DNA from CEM cells harvested 9 days after infection. Unintegrated HIV-1 DNA was analysed by nuclease S1 treatment either with no previous restriction enzyme digestion (lanes 1, 2, 5, and 6) or after digestion with *Bam*HI (lanes 3, 4, 7, and 8). Lanes 1, 3, 5, and 7, No nuclease S1 treatment; lanes 2, 4, 6, and 8, treatment with nuclease S1 (1.5 U/μg of DNA). Both Southern blots were hybridized to the 3′ probe shown in Fig. 3B. Both autoradiograms are overnight exposures.

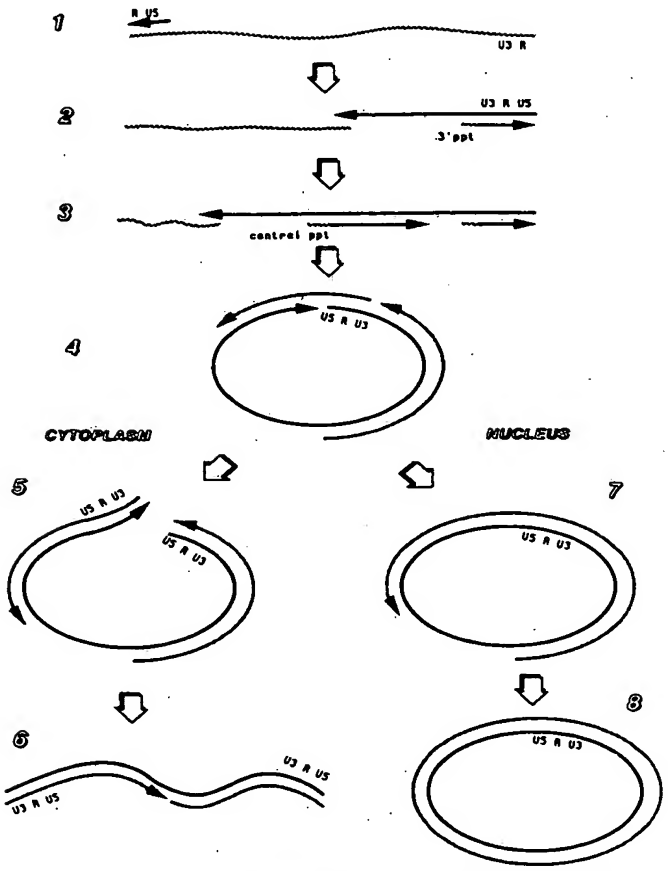


FIG. 6. Model for HIV reverse transcription. 1, Minus strong-stop synthesis; 2, first (minus-strand) template transfer and plus-strand strong-stop synthesis, initiated at the 3′ PPT; 3, progression of minus-strand synthesis and initiation of plus-strand synthesis at the central PPT; 4, second (plus-strand) template transfer; 5, formation of the LTRs by strand displacement and synthesis and progression of synthesis of the 5′ half of the plus strand; 6, linear gapped DNA molecule; 7, following step 4, ligation at both boundaries of the LTR, before strand displacement and synthesis can start, and progression of plus-strand synthesis; 8, ligation at the gap and formation of a one-LTR closed circle.

stranded gap are discrete. We have established that the 3′ boundary of this single-stranded element coincides with a PPT found at position 4800 in the HIV genome, an exact repeat of the structure normally found in all retroviruses next to the U3 region of the LTR (at position 9070 in the HIV-1 provirus). We have not determined the 5′ boundary of the gap and therefore cannot be certain of its size. However, analysis of the sizes of fragments released from nuclease S1-treated linear molecules following restriction enzyme digestions and corresponding to either the 3′ or the 5′ side of the gap suggests that this gap is short, probably less than 100 nucleotides in length.

In other retroviral models, the PPT has been shown to determine in vitro the initiation site of retroviral DNA plus-strand synthesis (20, 23, 27, 32). The DNA genomes of hepatitis B and cauliflower mosaic viruses, which are synthesized through reverse transcription of an RNA template, carry several single-stranded structures. In particular, there are two plus-strand discontinuities in the cauliflower mosaic virus genome that are defined by short PPTs which likely correspond to plus-strand initiation sites (25, 42). Our findings indicate that in HIV-1, and most likely also in other lentiviruses, the central PPT is used in vivo as an additional priming site for plus-strand synthesis. Indeed, primer extension experiments shown here reveal that the plus-strand DNA 3′ to the gap starts exactly at the last nucleotide of this PPT. This finding establishes that this structure determines precise and specific priming of plus-strand DNA at a central position in the genome. Because we used *Taq* polymerase and not reverse transcriptase in the primer extension experiments, we could not conclude whether the RNA primer corresponding to the central PPT remained attached to the nascent DNA strand, as described for other retroviruses at the 3′ PPT (7, 26).

It has been shown in avian retroviruses that plus-strand synthesis can start at sites distinct from the 3′ PPT, resulting in plus-strand discontinuities (16, 18). However, these initiation sites do not seem to be unique or well defined, and they do not result in a discrete single-stranded structure comparable to what we describe here. It is likely that the resulting plus-strand segments are eliminated by strand displacement events, as was shown in mellitin-permeabilized virions (4). In HIV, such a strand displacement is likely to occur at the 3′ PPT, to generate the 5′ end of linear molecules (Fig. 6, steps 4 and 5), but seems not to occur at the central PPT. We cannot explain why the upstream limit of the gap remains discrete. The gapped linear molecules represent a defined species: they are full-length, double-stranded molecules on which synthesis of the 5′ half of the viral plus strand is

stopped near the initiation site of the 3′ half. It is possible, however, that similar to what is found in cauliflower mosaic virus (25), elongation of the upstream HIV plus strand engages in a brief strand displacement event. Further studies are needed to clarify this point.

As a whole, our findings support a model of retroviral DNA synthesis (Fig. 6) in which linear molecules require a strand displacement step for LTR synthesis (4, 26), whereas one-LTR circles are the results of ligation events. Indeed, closed circular DNA molecules, which are found exclusively in the nucleus (17, 31), could be generated by ligation at both boundaries of the LTR and at the gap (Fig. 6, steps 7 and 8), following the proposed "intrastrand" plus-strand template transfer (24). This, in turn, could result from early nuclear transport of uncompleted DNA molecules.

Finally, it remains to be understood why lentiviruses, unlike other retroviruses, have developed and conserved a repeat of the PPT at the center of the genome. The most likely explanation is that it allows progression of plus-strand synthesis before elongation of the minus strand is complete (Fig. 6, step 3), probably resulting in a gain of time in DNA synthesis. The precise location of the PPT repeat at the center of the genome supports this hypothesis: an additional plus-strand initiation at this site could allow its elongation to reach the 3′ PPT approximately at the same time minus-strand synthesis is completed, the latest being required for plus-strand template transfer (Fig. 6, step 4). The LTRs can then be synthesized as plus-strand synthesis is being completed in the 5′ half of the genome (Fig. 6, step 5). Since lentiviruses, which are not transforming viruses, rely essentially on reverse transcription for their propagation, this feature could constitute an evolutionary advantage. The initiation of HIV plus-strand synthesis at the center of the genome is also interesting in view of recent observations that in unstimulated normal human lymphocytes, synthesis of the HIV minus strand appears to be arrested approximately halfway along the genomic RNA template (44). Dependence of full-length DNA synthesis on cell growth had also been described for avian viruses (9, 40). If the described stop in HIV minus-strand synthesis is located beyond the central PPT, it may allow early synthesis of a plus strand covering the whole 3′ half of the genome and may direct faster completion of double-stranded full-length DNA molecules upon further lymphocyte activation.

ACKNOWLEDGMENTS

We thank Luc Montagnier, in whose laboratory these studies were conducted, Marc Alizon, Olivier Danos, and Dominic Dwyer for helpful comments during preparation of the manuscript, and Agnès Cordonnier, who gave us some of the DNAs analyzed here.

P.C. is a recipient of a grant from the Agence Nationale de Recherches sur le SIDA.

REFERENCES

1. **Adachi, A., H. E. Gendelmann, S. Koenig, T. Folks, R. Willey, A. Rabson, and M. A. Martin.** 1986. Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone. J. Virol. 59:284–291.
2. **Barré-Sinoussi, F., J. C. Chermann, F. Rey, M. T. Nugeyre, S. Chamaret, J. Gruest, C. Dauguet, C. Axler-Blin, F. Brun-Vezinet, C. Rouzioux, W. Rozenbaum, and L. Montagnier.** 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science 220:868–871.
3. **Blum, H. E., J. D. Harris, P. Ventura, D. Walker, K. Staskus, E. Retzel, and A. T. Haase.** 1985. Synthesis in cell culture of the gapped linear duplex DNA of the slow virus visna. Virology 142:270–277.
4. **Boone, L. R., and A. M. Skalka.** 1981. Viral DNA synthesized in vitro by avian retrovirus particles permeabilized with mellitin. II. Evidence for a strand displacement mechanism in plus-strand synthesis. J. Virol. 37:117–126.
5. **Chakrabarti, L., M. Guyader, M. Alizon, M. D. Daniel, R. C. Desrosiers, P. Tiollais, and P. Sonigo.** 1987. Molecular cloning and nucleotide sequence of simian immunodeficiency virus from macaques. Nature (London) 328:543–547.
6. **Feinberg, A. P., and B. Vogelstein.** 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132:6–13.
7. **Finston, W. I., and J. J. Champoux.** 1984. RNA-primed initiation of Moloney murine leukemia virus plus strands by reverse transcriptase in vitro. J. Virol. 51:26–33.
8. **Friedrich, R., and K. Moelling.** 1979. Effect of viral RNAse H on the avian sarcoma viral genome during early transcription in vitro. J. Virol. 31:630–638.
9. **Fritsch, E. M., and H. M. Temin.** 1977. Inhibition of viral DNA synthesis in stationary chicken embryo fibroblasts infected with avian retroviruses. J. Virol. 24:461–469.
10. **Gilboa, E., S. Mitra, S. Goff, and D. Baltimore.** 1979. A detailed model of reverse transcription and tests of crucial aspects. Cell 18:93–100.
11. **Grandgenett, D. P., G. F. Gerard, and M. Green.** 1972. Ribonuclease H: a ubiquitous activity in virions of ribonucleic acid tumor viruses. J. Virol. 10:1136–1142.
12. **Guyader, M., M. Emerman, P. Sonigo, F. Clavel, L. Montagnier, and M. Alizon.** 1987. Genetic organization and transactivation of human immunodeficiency virus type 2. Nature (London) 326: 662–669.
13. **Harada, S., Y. Koyanagi, and N. Yamamoto.** 1985. Infection of HTLV-III/LAV in HTLV-I-carrying cells MT-2 and MT-4 and application in a plaque assay. Science 229:568–566.
14. **Harris, J. D., J. V. Scott, B. Traynor, M. Brahic, L. Stowring, P. Ventura, A. T. Haase, and R. Peluso.** 1981. Visna virus DNA: discovery of a novel gapped structure. Virology 113:573–583.
15. **Hirt, B.** 1967. Selective extraction of polyoma DNA from infected mouse cell cultures. J. Mol. Biol. 26:365–369.
16. **Hsu, T. W., and J. M. Taylor.** 1982. Single-stranded regions on unintegrated avian retrovirus DNA. J. Virol. 44:47–53.
17. **Kim, S., R. Byrn, J. Groopman, and D. Baltimore.** 1989. Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: evidence for differential gene expression. J. Virol. 63:3708–3713.
18. **Kung, H. J., Y. K. Fung, J. E. Majors, J. M. Bishop, and H. E. Varmus.** 1981. Synthesis of plus strands of retroviral DNA in cells infected with avian sarcoma virus and mouse mammary tumor virus. J. Virol. 37:127–138.
19. **Méchali, M., A. M. de Recondo, and M. Girard.** 1973. Action of the S1 endonuclease from Aspergillus oryzae on simian virus 40 supercoiled component I DNA. Biochem. Biophys. Res. Commun. 54:1306–1320.
20. **Mitra, S., M. Chow, J. J. Champoux, and D. Baltimore.** 1982. Synthesis of murine leukemia virus plus strong stop initiates at a unique site. J. Biol. Chem. 257:5983–5986.
21. **Mitra, S., S. Goff, E. Gilboa, and D. Baltimore.** 1979. Synthesis of a 600 nucleotide-long plus-strand DNA by virions of Moloney murine leukemia virus. Proc. Natl. Acad. Sci. USA 76:4355–4359.
22. **Olsen, J., and K. Watson.** 1982. Reverse transcription of avian myeloblastosis 35S RNA. Early synthesis of plus-strand DNA of discrete size in reconstructed reactions. Nucleic Acids Res. 10:1009–1027.
23. **Omer, C. A., R. Resnick, and A. J. Faras.** 1984. Evidence for involvement of an RNA primer in initiation of strong-stop plus-strand DNA synthesis during reverse transcription. J. Virol. 50:465–470.
24. **Panganiban, A. T., and D. Fiore.** 1988. Ordered interstrand and intrastrand DNA transfer during reverse transcription. Science 241:1064–1069.

24a. **Peden, K.** Unpublished data.

25. **Pfeiffer, P., and T. Hohn.** 1983. Involvement of reverse transcription in the replication of cauliflower mosaic virus: a detailed model and test of some aspects. Cell 31:781–789.
26. **Rattray, A. J., and J. J. Champoux.** 1987. The role of Moloney murine leukemia virus RNase H activity in the formation of plus-strand primers. J. Virol. 61:2843–2851.
27. **Resnick, R., C. A. Omer, and A. J. Faras.** 1984. Involvement of retrovirus reverse transcriptase-associated RNase H in the initiation of strong-stop (+) DNA synthesis and the generation of the long terminal repeat. J. Virol. 51:813–821.
28. **Rey, M. A., B. Krust, A. G. Laurent, L. Montagnier, and A. G. Hovanessian.** 1989. Characterization of human immunodeficiency virus type 2 envelope glycoproteins: dimerization of the glycoprotein precursor during processing. J. Virol. 63:647–658.
29. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463–5468.
30. **Shank, P. R., S. H. Hughes, H. J. Kung, M. J. E., L. Quintrell, R. V. Guntaka, J. M. Bishop, and H. E. Varmus.** 1978. Mapping unintegrated avian sarcoma virus DNA: termini of linear DNA bear 300 nucleotides present once or twice in two species of circular DNA. Cell 15:1383–1395.
31. **Shank, P. R., and H. E. Varmus.** 1978. Virus-specific DNA in the cytoplasm of avian sarcoma virus-infected cells is a precursor to covalently closed circular viral DNA in the nucleus. J. Virol. 25:104–114.
32. **Smith, J. K., A. Cywinski, and J. M. Taylor.** 1984. Initiation of plus-strand DNA synthesis during reverse transcription of an avian retrovirus genome. J. Virol. 49:200–204.
33. **Sonigo, P., M. Alizon, K. Staskus, D. Klatzmann, S. Cole, O. Danos, R. E., P. Tiollais, A. Haase, and S. Wain-Hobson.** 1985. Nucleotide sequence of the visna lentivirus: relationship to the AIDS virus. Cell 43:369–382.
34. **Southern, E.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98: 503–517.
35. **Staskus, K. A., M. S. Collett, and A. J. Faras.** 1976. Initiation of DNA synthesis by the avian oncornavirus RNA directed DNA polymerase: structural and functional localization of the major species of primer RNA on the oncornavirus genome. Virology 71:162–168.
36. **Taylor, J. M., and R. Ilmensee.** 1975. Site on the RNA of an avian sarcoma virus at which primer is bound. J. Virol. 16:553–558.
37. **Temin, H. M.** 1981. Structure, variation and synthesis of retrovirus long terminal repeats. Cell 27:1–3.
38. **Varmus, H., and R. Swanstrom.** 1982. Replication of retroviruses, p. 369–512. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses, vol. 1. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
39. **Varmus, H. E., S. Heasley, H. J. Kung, H. Oppermann, V. Smith, J. M. Bishop, and P. Shank.** 1978. Kinetics of synthesis, structure and purification of avian sarcoma virus-specific DNA made in the cytoplasm of acutely infected cells. J. Mol. Biol. 120:55–82.
40. **Varmus, H. E., T. Padgett, S. Heasley, G. Simon, and J. M. Bishop.** 1977. Cellular functions are required for the synthesis and integration of avian sarcoma virus-specific DNA. Cell 11:307–319.
41. **Verma, I. M.** 1975. Studies on reverse transcriptase of RNA tumor viruses. III. properties of purified Moloney murine leukemia virus DNA polymerase and associated RNase H. J. Virol. 15:843–854.
42. **Volovitch, M., G. Grugeon, and P. Yot.** 1978. Studies on the single strand discontinuities in the cauliflower mosaic virus genome. Nucleic Acids Res. 5:2913–2925.
43. **Wain-Hobson, S., P. Sonigo, O. Danos, S. Cole, and M. Alizon.** 1985. Nucleotide sequence of the AIDS virus, LAV. Cell 40:9–17.
44. **Zack, J. A., S. J. Arrigo, S. R. Weitzman, A. S. Go, A. Haislip, and I. S. Y. Chen.** 1990. HIV-1 entry into quiescent primary lymphocytes: molecular analysis reveals a labile, latent viral structure. Cell 61:213–222.